

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2022年1月13日(13.01.2022)



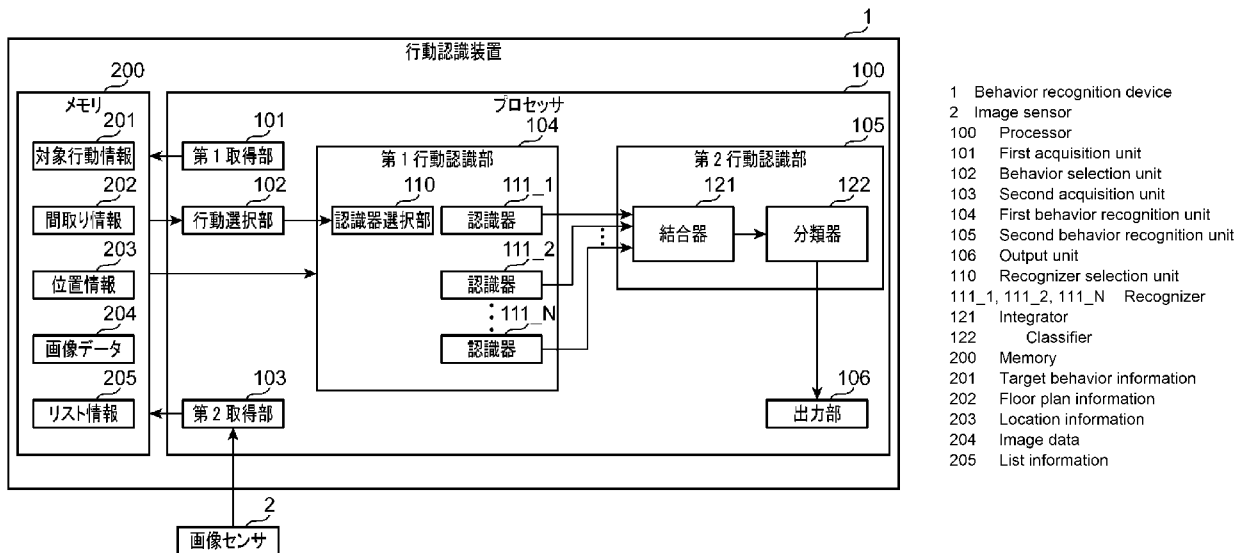
(10) 国際公開番号

WO 2022/009489 A1

- (51) 国際特許分類:
G06T 7/00 (2017.01) G06T 7/20 (2017.01)
- (21) 国際出願番号: PCT/JP2021/014244
- (22) 国際出願日: 2021年4月1日(01.04.2021)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2020-119202 2020年7月10日(10.07.2020) JP
- (71) 出願人: パナソニック インテレクチュアル プロパティ コーポレーション オブ アメリカ(PANASONIC INTELLECTUAL PROPERTY CORPORATION OF AMERICA) [US/US]; 90504 カリフォルニア州, トーランス, スイート 450, ウェスト 190 ストリート 2050 California (US).
- (72) 発明者: 若井 信彦 (WAKAI, Nobuhiko); 〒5718501 大阪府門真市大字門真1006番地 パナソニック株式会社内 Osaka (JP). 小塚 和紀(KOZUKA, Kazuki). 飯田 恵大(IIDA, Shigehiro). 中田 洋平(NAKATA, Yohei). 篠原 由美子(SHINOHARA, Yumiko).
- (74) 代理人: 小谷 昌崇, 外 (KOTANI, Masataka et al.); 〒5300005 大阪府大阪市北区中之島2丁目2番2号大阪中之島ビル2階 Osaka (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ,

(54) Title: BEHAVIOR RECOGNITION DEVICE, BEHAVIOR RECOGNITION METHOD, AND PROGRAM

(54) 発明の名称: 行動認識装置、行動認識方法、及びプログラム



(57) Abstract: This behavior recognition device selects a candidate behavior to be a recognition candidate from behaviors of interest on the basis of location information of an image sensor and floor plan information, acquires image data detected by the image sensor, determines one or more recognizers corresponding to the candidate behavior, calculates a feature amount of the image data using the one or more recognizers, and recognizes the candidate behavior on the basis of the feature amount.

(57) 要約: 行動認識装置は、画像センサの位置情報と間取り情報とに基づいて対象行動の中から認識候補となる候補行動を選択し、画像センサが検出した画像データを取得し、候補行動に応じた1以上の認識器を決定し、1以上の認識器を用いて画像データの特徴量を算出し、特徴量に基づいて候補行動を認識する。



NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT,
QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類：

- 一 国際調査報告 (条約第21条(3))

明 細 書

発明の名称： 行動認識装置、行動認識方法、及びプログラム

技術分野

[0001] 本開示は建物におけるユーザの行動を認識する技術に関するものである。

背景技術

[0002] 近年、動画像から人の行動を認識する研究が進められている。例えば、特許文献 1 には、動画像から人を含む人領域を検出し、前記領域に映る人の姿勢と周辺の物体との組み合わせから人の行動を認識する技術が開示されている。

[0003] 例えば、特許文献 2 には、動画像から人の関節に基づくスケルトン情報を時系列で抽出し、スケルトン情報の囲み領域を抽出し、抽出した囲み領域から人の行動を認識する技術が開示されている。

[0004] しかしながら、特許文献 1 及び特許文献 2 の技術によれば、建物の間取り情報が考慮されていないため、建物内の空間に依存する人の行動を精度よく認識するためには改善の必要があった。

先行技術文献

特許文献

[0005] 特許文献1： 2 0 1 8 - 2 0 6 3 2 1 号公報

特許文献2： 2 0 1 9 - 1 4 4 8 3 0 号公報

発明の概要

[0006] 本開示の一態様に係る行動認識装置は、建物における人の行動を認識する行動認識装置であって、予め定められた認識対象となる 1 以上の対象行動を含む対象行動情報と、前記建物の間取り情報と、前記建物に設置された画像センサの位置情報とを取得する第 1 取得部と、前記画像センサの位置情報と前記間取り情報とに基づいて前記対象行動情報に含まれる前記 1 以上の対象行動の中から認識候補となる候補行動を選択する行動選択部と、前記画像センサが検出した画像データを取得する第 2 取得部と、前記候補行動に応じた

1以上の認識器を決定し、前記1以上の認識器を用いて前記画像データの特徴量を算出する第1行動認識部と、前記特徴量に基づいて前記候補行動を認識する第2行動認識部と、前記第2行動認識部による認識結果を出力する出力部とを備える。

[0007] 本開示によれば、建物内の空間に依存する人の行動を精度よく認識することができる。

図面の簡単な説明

[0008] [図1]実施の形態に係る行動認識装置の構成の一例を示すブロック図である。

[図2]認識器を構成するCNNの説明図である。

[図3]第1行動認識部の構成の一例を示す図である。

[図4]行動選択テーブルの構成の一例を示す図である。

[図5]認識器選択テーブルの構成の一例を示す図である。

[図6]結合器が特徴量に対して重み係数を設定する際に参照される重みテーブルの一例を示す図である。

[図7]実施の形態に係る行動認識装置におけるリスト情報の生成処理の一例を示すフローチャートである。

[図8]図7のステップS106の処理の詳細を示すフローチャートである。

[図9]行動認識装置における行動認識処理の一例を示すフローチャートである。

[図10]認識器に関する情報を纏めた事前知識テーブルのデータ構成を示す図である。

[図11]既存の認識器を個別に実行した場合の1フレームあたりの処理時間を纏めたテーブルである。

[図12]入力画像データを纏めた表である。

[図13]行動認識装置の認識精度を評価するために行われたシミュレーションの結果を纏めた表である。

[図14]本開示の変形例に係る行動認識装置の構成の一例を示すブロック図である。

[図15]本開示の変形例（３）に係る行動認識装置の構成の一例を示すブロック図である。

[図16]玄関に画像センサが設置されるシーンを示す図である。

[図17]玄関に画像センサが設置されるシーンにおいてユーザと表示端末とのインタラクションの一例を示す図である。

[図18]図１７に続くインタラクションを示す図である。

[図19]アノテーション画像が重畳された設定画面の一例を示す図である。

[図20]キッチンに画像センサが設置されるシーンを示す図である。

[図21]キッチンに画像センサが設置されるシーンにおいてユーザと表示端末とのインタラクションの一例を示す図である。

[図22]図２１に続くインタラクションを示す図である。

[図23]アノテーション画像が重畳された設定画面の一例を示す図である。

[図24]画像センサの設置後にアノテーション画像を修正する際のユーザと表示端末とのインタラクションの一例を示す図である。

[図25]アノテーション画像が重畳表示された設定画面の一例を示す図である。

発明を実施するための形態

[0009] （本開示に至る知見）

従来、センサデータである動画像または静止画像から人の行動を推定する方法が提案されている。このような方法においては、認識すべき対象の行動に対して予めラベルが付与され、センサデータがどのラベルに該当するかを判定することが行われる。センサデータが動画像または静止画像である場合、畳み込み層及びプーリング層を用いたディープニューラルネットワーク（DNN）を使用することで高い認識精度が実現可能である。しかしながら、現状のDNNを使用した方法では、認識対象とする行動が人の歩行のような動作する空間に依存しない行動が認識対象とされている。そのため、空間に依存する行動を認識対象とするためには空間に関する情報を考慮に入れてDNNを一から学習させることが要求され、手間及びコストがかかる。

[0010] 上述の特許文献1では、人の姿勢と人の周辺の物体との組み合わせを考慮して人の行動が認識されているが、人の周囲に、人の行動とは無関係な物体が映り込んでいる場合、その物体は人の行動を推定するに際して有用な情報にはならない。そのため、特許文献1では、空間に依存する人の行動を認識するにはさらなる改善が必要である。

[0011] 上述の特許文献2では、スケルトン情報の囲み領域から人の行動が推定されていることから明らかなように、認識対象の行動として背景に依存しない行動が想定されている。そのため、特許文献2では、例えばキッチンで行う調理行動のように、空間に強く依存する行動を推定するにはさらなる改善が必要である。

[0012] そこで、本発明者は、建物の間取り情報を考慮すれば、建物内の空間に依存する人の行動を精度よく認識できるとの知見を得て、以下に示す本開示の各態様を想到するに至った。

[0013] 本開示の一態様に係る行動認識装置は、建物における人の行動を認識する行動認識装置であって、予め定められた認識対象となる1以上の対象行動を含む対象行動情報と、前記建物の間取り情報と、前記建物に設置された画像センサの位置情報とを取得する第1取得部と、前記画像センサの位置情報と前記間取り情報とに基づいて前記対象行動情報に含まれる前記1以上の対象行動の中から認識候補となる候補行動を選択する行動選択部と、前記画像センサが検出した画像データを取得する第2取得部と、前記候補行動に応じた1以上の認識器を決定し、前記1以上の認識器を用いて前記画像データの特徴量を算出する第1行動認識部と、前記特徴量に基づいて前記候補行動を認識する第2行動認識部と、前記第2行動認識部による認識結果を出力する出力部とを備える。

[0014] 本構成によれば、画像センサの位置情報と建物の間取り情報とに基づいて対象行動の中から認識対象となる候補行動が選択され、候補行動に対する認識結果が算出されている。そのため、建物内の空間に依存する人の行動を精度よく認識することが可能となる。

- [0015] さらに、本構成によれば、第1行動認識部が候補行動に応じた認識器を決定し、決定した認識器を用いて画像データの特徴量を算出し、算出された特徴量から第2行動認識部が候補行動を認識している。そのため、認識器として既存の認識器を利用することが可能となり、行動認識装置の構築が容易になる。さらに、候補行動に応じた1以上の認識器を用いて特徴量が算出されているため、候補行動の認識に適した特徴量を算出することが可能となり、対象行動の認識精度を高めることが可能となる。
- [0016] 上記行動認識装置において、前記第1行動認識部は、前記候補行動が所定の行動である場合、複数の認識器を決定し、前記第2行動認識部は、前記複数の認識器がそれぞれ算出した特徴量を結合し、結合した特徴量に基づいて前記候補行動を認識してもよい。
- [0017] 本構成によれば、対象行動が所定の行動である場合、複数の認識器が決定されているため、例えば所定の行動が人の周囲にある物体に依存するような場合において、人の特徴量を算出する認識器と、物体の特徴量を算出する認識器とを用いて特徴量を算出することが可能となり、候補行動の認識精度を高めることが可能となる。
- [0018] さらに、本構成によれば、各認識器が算出した特徴量が結合され、結合された特徴量に基づいて対象行動が認識されている。そのため、複数の認識器により複数の特徴量が算出された場合において、1つの特徴量に結合して第2行動認識部に入力することが可能となり、行動認識装置の構成を簡素化することが可能となる。
- [0019] 上記行動認識装置において、前記所定の行動は、掃除、歯磨き、料理、洗濯、コンピュータの使用、読書、又は食事であってもよい。
- [0020] 本構成によれば、掃除又は歯磨きといった人の行動が人の周囲の物体に依存するような行動の特徴量を画像データから精度よく算出することができる。
- [0021] 上記行動認識装置において、各認識器は、コンボリユーションニューラルネットワーク（以下、CNNと呼ぶ。）で構成され、前記第2行動認識部は

、ロジスティクス回帰、サポートベクターマシン、決定木、ランダムフォレスト、k近傍法、ガウシアンナイーブベイズ、パーセプトロン、及び確率的降下法のいずれか1つを用いた分類器を用いて前記候補行動を認識してもよい。

[0022] 本構成によれば、各認識器はCNNで構成され、第2行動認識器はロジスティック回帰等を用いた分類器により構成されているため、第1行動認識部に対して処理コストが低い分類器で第2行動認識器を構成することが可能となる。

[0023] 上記行動認識装置において、前記第2行動認識部は、前記特徴量を説明変数とし、前記対象行動を目的変数として機械学習された分類器を用いて前記候補行動を認識してもよい。

[0024] 本構成によれば、各認識器が算出した特徴量を説明変数とし、その特徴量に対応する候補行動を目的変数とする機械学習により生成された分類器を用いて第2行動認識部を構成することができる。例えば、分類器を上述したロジスティック回帰等のCNNに比べて処理負荷が小さい分類器で構成すれば、短時間で分類器を学習させることができる。さらに、第1行動認識部をCNNで構成された既存の認識器で構成することにより、第2行動認識部のみ機械学習させ、第1行動認識部を機械学習させることなく行動認識装置を構成できる。

[0025] 上記行動認識装置において、前記第2行動認識部は、前記候補行動に応じて予め定められた重み係数を用いて各特徴量に重み付けをし、重み付け後の各特徴量に基づいて前記候補行動を認識してもよい。

[0026] 本構成によれば、対象行動に応じて各特徴量に重み付けが行われ、重み付けされた各特徴量に基づいて候補行動が認識されるため、候補行動を正確に認識できる。

[0027] 上記行動認識装置において、前記間取り情報から、前記建物に設置された1以上の物体を抽出し、前記1以上の物体を、移動可能な第1物体、水回り設備である第2物体、及び前記建物の構造物である第3物体のいずれかに分

類し、前記 1 以上の物体のそれぞれについて、分類結果を示す分類情報と設置位置とが対応付けられた間取り特徴量を抽出し、前記間取り特徴量に基づいて、行動選択テーブルを生成し、前記行動選択テーブルは、前記建物の 1 以上の空間と、各空間に対応する前記対象行動とが対応付けられたテーブルであり、前記第 1 取得部は、前記行動選択テーブルを前記対象行動情報として取得してもよい。

[0028] 本構成によれば、間取り情報に基づいて、建物に設置された各物体について、分類情報と設置位置とが対応付けられた間取り特徴量が抽出される。そのため、建物内の各空間においてどのような物体が設置されているかが把握することが可能となり、行動選択テーブルを作成する上で有用な情報を抽出することができる。そして、間取り特徴量に基づいて、建物の各空間と、各空間に対応する対象行動とが対応付けられた行動選択テーブルが生成され、その行動選択テーブルが対象行動情報として取得される。そのため、各空間と対象行動との関係性を速やかに把握することができ、候補行動を容易に選択することができる。

[0029] 上記行動認識装置において、前記行動認識装置は表示端末と通信可能に接続され、前記画像センサが設置される前記建物の空間の名称を前記表示端末を介して取得し、前記空間に関連する特定機器又は特定設備が前記画像センサの視野内に含まれるように前記画像センサを設置する設置ガイダンスを表示端末に出力する設置支援部をさらに備えてもよい。

[0030] 本構成によれば、空間に関連する特定機器又は特定設備が画像センサの視野内に含まれるように前記画像センサを設置する設置ガイダンスが表示端末に出力されるため、ユーザは画像センサを適切に設置できる。さらに、建物の空間の名称が表示端末を介して取得されているため、画像センサと画像センサが設置された空間とを対応付けることができる。その結果、画像センサの設置位置と行動選択テーブルとを参照して、候補行動を容易に選択できる。

[0031] 上記行動認識装置において、前記設置支援部は、前記画像センサが撮影し

た画像データを取得し、前記画像データに含まれる前記特定機器又は前記特定設備を検出し、前記特定機器又は前記特定設備の検出結果を示すアノテーション画像を、前記画像データが示す画像に重畳表示させてもよい。

[0032] 本構成によれば、画像データが示す画像に、画像データから検出された特定機器又は特定設備の検出結果であるアノテーション画像が重畳表示されている。そのため、画像データから特定機器又は特定設備が正しく検出されているかを容易に確認することができる。

[0033] 上記行動認識装置において、前記設置支援部は、前記アノテーション画像の修正指示を前記表示端末を介して取得し、修正された前記アノテーション画像が示すアノテーション情報をメモリに記憶してもよい。

[0034] 本構成によれば、アノテーション情報の修正指示が表示端末を介して取得されるため、例えば認識器が特定機器又は特定設備を正しく検出できなかった場合、特定機器又は特定設備の位置を正確に示すようにアノテーション情報を修正できる。さらに、認識器に対して対象行動を認識する上で重要となる特定機器又は特定設備の画像上の位置を把握させることができる。

[0035] 本開示は、このような行動認識装置に含まれる特徴的な各構成をコンピュータが実行する行動認識方法及び特徴的な各構成をコンピュータに実行させる行動認識プログラムとして実現することもできる。また、このような行動認識プログラムを、CD-ROM等のコンピュータ読取可能な非一時的な記録媒体或いはインターネット等の通信ネットワークを介して流通させることができるのは、言うまでもない。

[0036] なお、以下で説明する実施の形態は、いずれも本開示の一具体例を示すものである。以下の実施の形態で示される数値、形状、構成要素、ステップ、ステップの順序などは、一例であり、本開示を限定する主旨ではない。また、以下の実施の形態における構成要素のうち、最上位概念を示す独立請求項に記載されていない構成要素については、任意の構成要素として説明される。また全ての実施の形態において、各々の内容を組み合わせることもできる。

[0037] (実施の形態)

コンピュータを用いた人の行動の認識手法は、人と人の居る空間との関係で3つのタイプに分類される。第1のタイプは、人が居る空間と無関係な行動を認識する手法である。例えば、歩く又は立つ等の行動は着目する人の動きのみで表現され、人の居る空間とは無関係である。第2のタイプは、人の居る空間に存在する物体又は人の近くに位置する物体と関与する行動を認識する手法である。例えば、自転車に乗っているという行動は、人の姿勢検出と物体検出とを組み合わせることで認識可能である。屋外等で物体に関与する人の行動を認識する場合、建物の構造又は道路の状態等、予め把握することが困難である。そのため、このような行動は、人と人の周辺にある物体とを利用した認識処理を行うことで、認識可能となる。第3のタイプは、人と人の居る空間との両方を把握することによって行動を認識する手法である。例えば、キッチンで料理を開始する行動は、人がキッチンに存在し、電子レンジの方向を向くことにより認識できる。

[0038] このように、人の居る空間の情報を利用しない場合に認識が困難となる「料理を開始する」といった行動は、従来の認識手法である第1のタイプ及び第2のタイプでは認識できない。本実施の形態では、人と人がいる空間の情報である間取り情報とを用いて、人の行動を認識する手法を開示する。以下、本実施の形態について、図面を参照しながら説明する。本実施の形態が認識対象とする人の行動は、例えば、料理を開始する、掃除をするといった動きを伴う行動のほか、例えば、テレビを寝ながら見るといった動きを伴わない行動も含む。

[0039] 図1は、実施の形態に係る行動認識装置1の構成の一例を示すブロック図である。行動認識装置1は、ユーザの宅内（建物の一例）においてユーザの行動を認識する装置である。

[0040] 行動認識装置1は、例えば、プロセッサ、メモリ、インタフェース回路などからなるコンピュータで構成されている。行動認識装置1は、単一のコンピュータで実現される必要はなく、端末装置とサーバとを含む分散処理シス

テム（不図示）によって実現されてもよい。例えば、行動認識装置１は、画像データ２０４を格納するメモリを宅内の端末装置に設け、プロセッサ１００を構成する一部又は全部のブロックをサーバに設けて構成されてもよい。この形態については、後述する変形例で説明する。

[0041] 行動認識装置１は、プロセッサ１００、及びメモリ２００を含む。メモリ２００は、ＳＳＤ又はＨＨＤ等の不揮発性の記憶装置で構成され、対象行動情報２０１、間取り情報２０２、位置情報２０３、画像データ２０４、及びリスト情報２０５を記憶する。メモリ２００は、第２取得部１０３が所定のフレームレートで画像センサ２から取得した画像データ２０４のうち現在から過去に遡って一定時間（例えば１分）の画像データ２０４を保存すればよい。

[0042] 画像センサ２は、例えば宅内に設置されたカメラで構成されている。画像センサ２は、所定のフレームレートで宅内の空間を撮影することで、画像データ２０４を取得し、第２取得部１０３に入力する。画像センサ２は複数であってもよい。画像データ２０４は、例えば、カラーの画像データであってもよいし、モノクロの画像データであってもよい。

[0043] プロセッサ１００は、ＣＰＵ等の電気回路で構成されている。プロセッサ１００は、第１取得部１０１、行動選択部１０２、第２取得部１０３、第１行動認識部１０４、第２行動認識部１０５、及び出力部１０６を含む。これらの各ブロックは、プロセッサ１００が例えば行動認識プログラムを実行することで実現される。但し、これは一例であり、これらの各ブロックはＡＳＩＣ等の専用のハードウェア回路で実現されてもよい。

[0044] 第１取得部１０１は、予め定められた認識対象となる対象行動を示す対象行動情報２０１を取得し、メモリ２００に格納する。例えば、第１取得部１０１は、図略の入力装置を用いた登録作業を通じて入力された対象行動情報２０１を取得すればよい。入力装置は例えばキーボード、及びマウス等で構成される。但し、これは一例であり、対象行動情報２０１は予めメモリ２００に記憶されていてもよい。対象行動は例えば図４に示す行動選択テーブル

T 1 に登録された拭き掃除、掃き掃除等の各種行動である。対象行動情報 201 は事前にメモリ 200 に格納されていてもよい。行動選択テーブル T 1 は対象行動情報 201 の一例である。

[0045] さらに、第 1 取得部 101 は、宅内の間取り情報 202 を取得し、メモリ 200 に格納する。間取り情報 202 は、リビング、ダイニング、キッチンといった部屋の構成要素と、各部屋の形状及び位置関係とを表現した 2 次元又は 3 次元の情報である。間取り情報 202 は、例えば間取りを記載する 2 次元の図面データ、住宅設計に使用される 3 次元設計データ（CAD データ）、3 次元レーザースキャナで計測された宅内の 3 次元点群情報（ポイントクラウド）、掃除ロボット等の宅内を移動する機器の軌跡情報、校正済みのカメラで撮影された宅内の映像データ、又は設置された部屋の情報を有する機器から取得された情報で構成される。

[0046] さらに、第 1 取得部 101 は、宅内に設置された画像センサ 2 の位置情報 203 を取得し、メモリ 200 に格納する。位置情報 203 は、例えば間取り情報 202 に含まれる 2 軸又は 3 軸の座標空間の座標系で表される座標データで構成される。第 1 取得部 101 は、例えば図略の入力装置を用いた登録作業によって入力された位置情報 203 を取得する。

[0047] 行動選択部 102 は、画像センサ 2 の位置情報 203 と間取り情報 202 とに基づいて対象行動の中から認識候補となる候補行動を選択する。例えば、行動選択部 102 は、位置情報 203 と間取り情報 202 とから画像センサ 2 が設置された宅内の空間を特定し、特定した空間においてユーザが行動することが推定される予め定められた行動を選択する。具体的には、行動選択部 102 は、特定した空間に対応する行動を図 4 に示す行動選択テーブル T 1 を参照することで決定すればよい。空間は、例えば部屋、玄関、キッチン等の宅内を構成する空間である。

[0048] 第 2 取得部 103 は、画像センサ 2 が所定のフレームレートで撮影した画像データ 204 を取得し、メモリ 200 に格納する。

[0049] 第 1 行動認識部 104 は、行動選択部 102 が選択した候補行動に応じた

1以上の認識器を決定し、決定した認識器を用いて画像データ204の特徴量を算出する。第1行動認識部104は、認識器選択部110、N（Nは1以上の整数）個の認識器111__1、111__2、・・・111__Nを含む。以下、認識器111__1、111__2、・・・111__Nを総称する場合、認識器111と表す。

[0050] 認識器選択部110は、行動選択部102が選択した候補行動の認識に用いられる認識器111を選択する。例えば認識器選択部110は、後述するリスト情報205の生成時には図5に示す認識器選択テーブルT2を参照することで、候補行動に対応する認識器111を選択し、認識処理の実行時には後述のリスト情報205を参照することで、候補行動に対する認識器111を選択すればよい。

[0051] 認識器111は、例えばCNNで構成された認識器である。本実施の形態では、第1行動認識部104は、図5に示すように、姿勢推定、物体検出、顔検出、頭向き推定、年齢性別推定、個人推定、及びトラッキングのそれぞれを個別に認識する認識器111を含む。これらの認識器111は、ソースコードが公開された既存の認識器を流用又は一部改変したものである。

[0052] 特徴量は認識器111に応じて異なる。姿勢推定の認識器111が算出する特徴量には、例えば骨格情報を構成する複数の特徴点（例えば、右肩、右肘のような17点）のそれぞれを表す2次元の座標データが含まれる。

[0053] 物体検出の認識器111が算出する特徴量には、例えば物体を取り囲む外接矩形の座標データ及び認識した物体のラベルが含まれる。物体検出で検出される物体は所与であり、例えば80種類の宅内の物体（電子レンジ及び冷蔵庫等）が検出される。顔検出の認識器111が算出する特徴量には、例えば顔を外接矩形で取り囲んだ顔領域の画像データが含まれる。ここで、各認識器111が算出する特徴量は、DNNの中間層から出力されるような人間が理解できないデータ（例えば、テンソル）ではない。

[0054] 第1行動認識部104は、候補行動に応じて他の認識器111に依存する認識器111を含む。依存する認識器111とは、他の認識器111が算出

した特徴量が入力され、その特徴量を利用して特徴量を算出する認識器 1 1 1 である。例えば、頭向き推定の認識器 1 1 1 は、顔検出の認識器 1 1 1 が算出した特徴量（顔領域の画像データ）を用いて頭の向きを示す特徴量を算出する。

[0055] 第 2 行動認識部 1 0 5 は、各認識器 1 1 1 が算出した特徴量に基づいて行動選択部 1 0 2 が選択した各候補行動を認識する。第 2 行動認識部 1 0 5 は、結合器 1 2 1 及び分類器 1 2 2 を含む。結合器 1 2 1 は、各結合器 1 2 1 が算出した特徴量を結合し、結合した特徴量を分類器 1 2 2 に入力する。ここで、結合器 1 2 1 は、候補行動に応じて予め定められた重み係数を用いて各特徴量に重み付けをし、重み付け後の各特徴量を連結して分類器 1 2 2 に入力してもよい。

[0056] 分類器 1 2 2 は、結合器 1 2 1 から入力された特徴量に基づいて候補行動に対する尤度を算出することによって候補行動を認識する。分類器 1 2 2 は、クラス分類を行う分類器で構成されている。分類器 1 2 2 は、例えばロジスティクス回帰、サポートベクターマシン、決定木、ランダムフォレスト、k 近傍法、ガウシアンナイーブベイズ、パーセプトロン、及び確率的降下法のいずれか 1 つを用いた分類器で構成されている。候補行動が複数の場合、分類器 1 2 2 は、各候補行動に対する尤度をそれぞれ算出する。

[0057] 分類器 1 2 2 は、認識器 1 1 1 から出力される特徴量を説明変数とし、対象行動を目的変数として機械学習されたものである。この機械学習は、各対象行動に対して個別に行われる。例えば、対象行動が「料理を開始する」の場合、この対象行動に使用される各認識器 1 1 1 から出力される特徴量の結合結果が説明変数とされ、「料理を開始する」が目的変数とされて機械学習が行われる。

[0058] 出力部 1 0 6 は、第 2 行動認識部 1 0 5 の認識結果を出力する。具体的には、出力部 1 0 6 は、分類器 1 2 2 が算出した尤度に基づいて人の行動に対する認識結果を出力する。例えば、候補行動が複数の場合、分類器 1 2 2 は、尤度が最大のラベルを認識結果として出力すればよい。

- [0059] 認識結果は、例えば通信回路を介して外部機器（図略）に出力される。外部機器は、例えば宅内に設置され、認識結果に基づいた制御を実行する家電機器で構成されてもよい。この家電機器は、例えば宅内に設置され、認識結果を表示する表示装置であってもよい。
- [0060] 図2は、認識器111を構成するCNNの説明図である。CNNは、畳み込み層及び全結合層を含む。この例では、CNNは畳み込み層が9層、全結合層が2層で構成されている。図2において、各種記号の添え字は入力画像データD1の入力層を第0層、畳み込み層を第1層～第9層、全結合層を第10層～第11、出力層を第12層としたときの層番号を示している。入力画像データD1は、 $H0 \times W0 \times C0$ で表される。ここで、Hは各層におけるデータの高さ、Wは各層におけるデータの幅、Cは各層におけるチャンネル数を示している。入力画像データD1がカラーの画像データの場合、 $C0 = 3$ となる。畳み込み層では、畳み込み演算により各層のデータの特徴量が算出される。例えば、第5層から出力されるデータのサイズ、すなわち、 $H5 \times W5 \times C5$ は、 $64 \times 64 \times 256 = 1048576$ で表される。
- [0061] 全結合層では、畳み込み層から出力されるデータがまとめられ、出力データが生成される。例えば、M点の骨格情報を最大P人検出する姿勢推定のCNNの場合、全結合層からは $P \times M \times 2 (x, y)$ 個のデータが出力データとして出力される。
- [0062] CNNは画像データに対する認識精度が非常に高い。しかし、CNNは畳み込み層及び全結合層というように数多くの層で構成されているため、処理負荷が高く学習には多大な時間がかかる。これに対して、分類器122は、CNNに比べて処理負荷が大幅に低いサポートベクターマシン等の分類器である。そこで、本実施の形態では、CNNから構成される認識器111として既存の認識器を利用し、処理負荷の低い分類器122に対象行動を学習させている。これにより、学習時間の短縮が図られ、行動認識装置1を容易に構築することが可能となる。
- [0063] 図3は、第1行動認識部104の構成の一例を示す図である。この例では

、第1行動認識部104は、姿勢推定を行う認識器111__1、物体検出を行う認識器111__2、顔検出を行う認識器111__3、及び頭向き推定を行う認識器111__4を含む。認識器111__1～111__3のそれぞれには入力画像データD1が入力されている。入力画像データD1は画像センサ2により撮影された宅内の画像データである。

[0064] 以下、対象行動が「料理を開始する」の場合を例に挙げて説明する。例えば、入力画像データD1は、キッチンで人が冷蔵庫の方を向いて立っているシーンを含む。認識器111__1は、人が起立している入力画像データD1が入力された場合、その起立した姿勢の骨格情報を構成する座標データを特徴量として出力する。認識器111__2は、入力画像データD1から所与の物体を検出し、検出した物体を囲む外接矩形の頂点の座標データと検出した物体のラベルとを特徴量として出力する。ここでは、物体として冷蔵庫が検出される。

[0065] 認識器111__3は、入力画像データD1から人の顔を検出し、検出した顔を外接矩形で囲む顔領域の画像データを特徴量として出力する。認識器111__4は、認識器111__3から出力される顔領域の画像データから人の頭向きを推定する。頭向きは、例えば顔領域の重心を起点とする方向ベクトルで表される。

[0066] 第2行動認識部105は、例えば方向ベクトルの方向に冷蔵庫を囲む外接矩形の重心がある場合、人が冷蔵庫を向いていると判定する。例えば、第2行動認識部105は、画像上において方向ベクトルを起点からプラスマイナス閾値角度（例えば、プラスマイナス30度）の範囲で回転させ、前記冷蔵庫の重心を前記方向ベクトルの延長線が通過する場合、人は冷蔵庫を向くと判定する。

[0067] これらの人の姿勢、物体位置、及び頭向きの情報を組み合わせることで、第2行動認識部105は、「料理を開始する」という行動を判定することができる。なお、認識器111__4は、頭向きの代わりに、視線の方向、鼻の方向、胴体の法線方向を検出してもよい。

[0068] 図4は、行動選択テーブルT1の構成の一例を示す図である。行動選択テーブルT1は、複数の空間と、各空間のそれぞれについて認識対象となる対象行動とを対応付けて記憶する。対象行動としては、拭き掃除、掃き掃除、歯磨き、料理を開始する、コーヒーを淹れる、ノートパソコン（ノートPC）の使用、洗濯、読書、食事、及び歩くが採用されている。空間としては、玄関、キッチン、リビング、ダイニング、ベッドルーム、バスルーム、及び洗面所が採用されている。但し、これらは一例であり、他の空間及び対象行動が採用されてもよい。例えば、ソファで横になりテレビを見ているといった対象行動がリビングに対して割り当てられてもよい。

[0069] 行動選択テーブルT1において、○マークは該当する空間に対して該当する対象行動が認識対象とされることが示されている。×マークは該当する空間に対して該当する対象行動が認識対象とされていないことを示している。

[0070] 例えば、1行目に示す拭き掃除は、玄関、キッチン、リビング、ダイニング、バスルーム、及び洗面所において認識対象となる。

[0071] 拭き掃除は全空間で行われる可能性が高いため、全空間において認識対象とされている。掃き掃除はバスルームで行われる可能性が低いため、バスルーム以外の空間で認識対象とされている。歯磨きは水回りの空間で行われる行動であるため、キッチン、バスルーム、及び洗面所において認識対象とされている。

[0072] 料理を開始するは、キッチン以外では行われなため、キッチンのみが対象空間とされている。コーヒーを淹れるは、コーヒーメーカーがキッチンのみならずダイニングにも置かれる可能性があるため、キッチン及びダイニングが認識対象とされている。

[0073] ノートPCの使用は、ノートPCがダイニング、リビング、及びベッドルームで使用される可能性があるため、これらの空間において認識対象とされている。

[0074] 洗濯は洗濯機が洗面所に設置されることを想定して、洗面所において認識対象とされている。読書は人が座る空間において行われる可能性が高いため

、リビング、ダイニング、及びベッドルームにおいて認識対象とされている。食事はダイニングのみならずリビングで行われる可能性もあるため、リビング及びダイニングにおいて認識対象とされている。歩くは全空間で行われるため、全空間において認識対象とされている。

[0075] 行動選択部 102 は、このような行動選択テーブル T1 を参照することで空間に対応する対象行動を選択できるため、空間に対して適切な対象行動を容易に選択することができる。

[0076] 図 5 は、認識器選択テーブル T2 の構成の一例を示す図である。認識器選択テーブル T2 は、複数の対象行動と、各対象行動を認識する際に使用される認識器 111 とを対応付けて記憶する。認識器選択テーブル T2 において、○マークは該当する対象行動に対して使用される認識器 111 を示し、×マークは該当する対象行動に対して使用されない認識器を示している。対象行動は、図 4 に示す対象行動と同じである。姿勢推定、物体検出、顔検出、頭向き推定、年齢性別推定、個人推定、及びトラッキングのそれぞれは、対応する認識器 111 により個別に実行される。

[0077] 例えば、拭き掃除は、個人を特定する必要がないため、顔検出、頭向き推定、年齢性別判定、及び個人推定の認識器 111 は使用されず、姿勢推定、物体検出、及びトラッキングの認識器 111 が使用される。

[0078] 掃き掃除は拭き掃除と同様の理由により、姿勢推定、物体検出、及びトラッキングの認識器 111 が使用される。歯磨きは、個人の特定及び移動が重要でないため、年齢性別推定、個人推定、及びトラッキング以外の認識器 111 が使用される。

[0079] 料理を開始するは、全ての認識器 111 が使用される。例えば、料理は母がする場合が多いことが想定されるため、料理を開始するでは、母か否かを推定するために年齢性別推定及び個人推定の認識器 111 が使用される。

[0080] コーヒーを淹れるは、料理を開始するとは異なり父又は子供が淹れることも想定され、移動は重要ではないので、年齢性別推定、個人推定、及びトラッキング以外の認識器 111 が使用される。

- [0081] ノートPCの使用は、PCを使用しながら人が動くことは稀であるため、トラッキング以外の認識器111が使用される。洗濯は、洗濯機が設置されている場所が固定されているため、トラッキング以外の認識器111が使用される。読書は、個人の特定が不要であり、動きもないため、年齢性別推定、個人推定、及びトラッキング以外の認識器111が使用される。
- [0082] 食事は、誰でも行い、頭向きは重要でなく、移動も想定されないため、姿勢推定及び物体検出の認識器111が使用される。歩くは、物体が関与せず、頭向き、及び個人の特定は重要ではないため、姿勢推定及びトラッキングの認識器111が使用される。上記以外の未知の行動を検出する場合、全ての認識器111が使用される。
- [0083] 次に、各対象行動の認識に使用される認識器111の有効性について説明する。
- [0084] 姿勢推定は、対象行動に応じた姿勢を検出することが可能であるため、全ての対象行動の認識に有効である。また、頭向き推定のためには、顔検出が必要である。つまり、頭向き検出は顔検出に依存する。
- [0085] 拭き掃除では、掃除機等の掃除道具の検出に物体検出が有効であり、複数の箇所を移動しながら掃除が行われるため、トラッキングが有効である。
- [0086] 掃き掃除では、拭き掃除と同様、物体検出及びトラッキングが有効である。
- [0087] 歯磨きでは、歯ブラシ及び鏡の検出に物体検出が有効であり、頭の向きが鏡の方向を向く傾向があるため、頭向き推定が有効である。
- [0088] 料理を開始するでは、調理器具及び食材の検出に物体検出が有効であり、頭の向きがキッチン又はシンクを向く傾向があるため頭向き推定が有効である。また、料理を開始するは、母が良く料理をするというように料理頻度が高い個人を特定するために年齢性別推定及び個人推定が有効であり、キッチンでシンクからコンロまで移動することがあるためトラッキングが有効である。
- [0089] コーヒーを淹れるでは、コーヒーマーカ及びコーヒーカップを検出するた

めに物体検出が有効であり、頭の向きがコーヒーマーカの方を向いているか否かを検出するために頭向き推定が有効である。

[0090] ノートP Cの使用ではノートP Cの検出に物体検出が有効であり、頭の向きがノートP Cの方を向いているか否かの検出に頭向き推定が有効であり、ノートP Cは仕事で使用することもあるため年齢性別及び個人推定が有効である。

[0091] 洗濯では、洗濯器の検出に物体検出が有効であり、頭の向きが洗濯器の方を向いているか否かの検出に頭向き推定が有効であり、母が良く洗濯をするというような個人を特定するために年齢性別推定及び個人推定が有効である。

[0092] 読書では、本の検出に物体検出が有効であり、頭の向きが本の方を向いているか否かの検出に頭向き推定が有効である。

[0093] 食事では、料理の検出に物体推定が有効である。

[0094] 歩くでは、連続した移動を捉えるためにトラッキングが有効である。

[0095] これらの対象行動以外においては、有効か否かは不明なため、全ての認識器111が選択される。

[0096] 行動認識装置1が認識対象とする対象行動は空間と関連している。例えば、「料理を開始する」は、キッチンという空間に強く関連しており、この空間の情報が行動認識に有用である。つまり、対象行動に対し、特定の物体又は頭向き等に対して重み付けをすれば、対象行動の検出精度がさらに向上する。そこで、本実施形態において、結合器121は、対象行動に応じた重み係数を特徴量に設定する。

[0097] 図6は、結合器121が特徴量に対して重み係数を設定する際に参照する重みテーブルT3の一例を示す図である。重みテーブルT3は、複数の対象行動と各対象行動に対するクラスインデックス、ラベル、空間情報、物体情報、頭向き情報、空間情報の重み係数 W_r 、物体情報の重み係数 W_e 、及び重み付けの内容とを対応付けて記憶する。ここでは、対象行動として、図4で例示した料理を開始するに加えて拭き掃除が例示されている。ここでは、

図示を省略しているが、実際には、図4で例示した各対象行動に関するレコードも重みテーブルT3には登録されている。

[0098] クラスインデックスは、各対象行動に対して一意に付与されたインデックスである。ここでは、クラスインデックスは連番で付与されている。ラベルは、出力部106が出力する対象行動のラベルである。ここでは、ラベルとして、「料理を開始する」及び「拭き掃除」が例示されている。ここでは、図示は省略しているが、重みテーブルT3には、図4で例示した各対象行動のラベルが登録されている。空間情報は対象行動が行われる空間の情報である。ここでは、「料理を開始する」に対してはキッチンというように、行動選択テーブルT1で規定された対象行動に対する空間が登録されている。

[0099] 物体情報は該当する対象行動に関連する物体を示す情報である。例えば、「料理を開始する」に対する物体情報は、冷蔵庫、電子レンジ、ガスコンロ、及びオーブンの少なくとも1つが該当する。頭向き情報は物体情報に記載した物体と頭向きとの関係を示す情報である。例えば、「料理を開始する」に対する頭向き情報は、冷蔵庫、電子レンジ、ガスコンロ、及びオーブンの少なくとも一つに頭が向くことである。これらの物体に対して頭が向いているか否かの判定は、上述した方向ベクトルと、物体の重心との関係に基づいて行われる。

[0100] 空間情報の重み係数 W_r は、対象行動が行われる空間に対する重み係数である。ここでは、行動選択テーブルT1において○マークが付された空間に対しては重み係数が「1」、×マークが付された空間に対しては重み係数が「0」に設定される。例えば、「料理を開始する」は、行動選択テーブルT1においてキッチンに○マークが付されているため、キッチンに対する重み係数 W_r は「1」、それ以外の空間に対する重み係数 W_r は「0」になる。

[0101] 物体情報の重み係数 W_e は、検出された物体に対する重み係数であり、該当する物体が検出された場合に「1」、それ以外の場合は「0」になる。一つの物体に対して複数の特徴量が抽出されている場合は、複数の特徴量のそれぞれに対して1または0が設定される。重み付けの内容は、検出された物

体と頭向きとの関係に応じて設定される重み係数の内容を示す。

[0102] 以下、図4で示す、姿勢推定、物体検出、及び頭向き推定の3つの認識器111から出力される特徴量を例に挙げて、結合器121による重み付けについて説明する。ここでは、結合器121が最終的に出力するベクトルを V_o とする。また、各認識器111が出力する特徴量もベクトルであるとする。なお、出力ベクトル V_o は一次元ベクトルに限らず、多次元のテンソルでも良い。重み付けをしない場合、出力ベクトル V_o は式(1)で表される。

[0103] [数1]

$$V_o = \langle V_p, V_{ob}, V_h \rangle \quad (1)$$

姿勢推定の認識器111、物体検出の認識器111、頭向き推定の認識器111が出力する特徴量をそれぞれ、 V_p 、 V_{ob} 、 V_h とする。重み付けがない場合の出力ベクトル V_o は、特徴量 V_p 、 V_{ob} 、 V_h を連結すればよい。出力ベクトル V_o の要素数は特徴量 V_p 、 V_{ob} 、 V_h の要素数の総和となる。記号 $\langle \rangle$ は $\langle \rangle$ 内のベクトルを、連結後もベクトルの方向に連結することを表す。

[0104] これに対して、重み付けをする場合、出力ベクトル V_o は式(2)で示される。重み付けをする場合、重み付けをしない場合と同様に、出力ベクトル V_o' は、各認識器111のベクトルを連結することで生成される。この際、各認識器111の特徴量に対して重み付けが行われる。したがって、出力ベクトル V_o' の要素数は重み付けがない場合と同じである。

[0105] [数2]

$$V_o' = W_r \odot \left(W_p \odot V_p, W_{ob} \odot (W_e \otimes V_{ob}), W_h \odot V_h \right) \quad (2)$$

式(2)において、 $\odot$ の中に「 $\cdot$ 」が付された記号はブロードキャスト演算を示し、 $\odot$ の中に $\times$ が付された記号はアダマール積を示す。

[0106] W_r 、 W_p 、 W_{ob} 、 W_h はそれぞれ部屋の重み係数、姿勢推定の重み係数、物体検出の重み係数、及び頭向き推定の重み係数である。重み係数 W_p

は特徴量 V_p に対してブロードキャスト演算され、重み係数 $W_o b$ は重み係数 W_e と特徴量 $V_o b$ とのアダマール積に対してブロードキャスト演算され、重み係数 W_h は特徴量 V_h に対してブロードキャスト演算される。ブロードキャスト演算は、重み係数（スカラー値）とベクトルの各要素との積を取る演算である。重み係数 W_e は物体の重み係数であり、特徴量 $V_o b$ と同じ要素数を持つ。

[0107] 例えば、対象行動が料理を開始するであり、入力画像データ D_1 がキッチンで撮影された画像データであり、特徴量 $V_o b$ に冷蔵庫のラベルが含まれ、特徴量 V_h に冷蔵庫を向く方向ベクトルが含まれていたとする。この場合、結合器 121 は、重みテーブル T_3 を参照し、重み係数 W_r を「1」、重み係数 W_e を「1」、重み係数 W_h を「5」、及び重み係数 $W_o b$ を「1」に設定する。

[0108] ここでは、説明を簡単にするため、対象行動が一つの場合を例に挙げて、式（2）の重み付けを説明した。対象行動が2つ以上の場合、結合器 121 は各対象行動に対して式（2）の重み付けを行い、重み付け後の特徴量を分類器 122 に入力すればよい。分類器 122 は、各対象行動に対して個別に尤度を算出すればよい。例えば、「料理を開始する」と「拭き掃除」との2つの行動を認識する場合、結合器 121 は、式（2）の各重み係数を「料理を開始する」に対応する値に設定し、次に、式（2）の重み係数を「拭き掃除」に対応する値に設定すればよい。

[0109] 対象行動が2つ以上の場合の別法として、例えば「料理を開始する」と「拭き掃除」のそれぞれの場合における重み係数の算術平均、算術和、又は論理和をとることで重み係数を統合する方法が採用可能である。

[0110] 図7は、実施の形態に係る行動認識装置1におけるリスト情報205の生成処理の一例を示すフローチャートである。リスト情報205の詳細は後述する。なお、図7のフローチャートは、例えば行動認識装置1が宅内に設置されたときに実行される。また、図7のフローチャートは、例えば間取り情報202又は位置情報203が変更されたときに実行されてもよい。

- [0111] まず、第1取得部101は行動選択テーブルT1（対象行動情報201）を取得してメモリ200に格納する（ステップS101）。次に、第1取得部101は間取り情報202を取得し、メモリ200に格納する（ステップS102）。
- [0112] 次に、第2取得部103は、画像センサ2の位置情報203を取得し、メモリ200に格納する（ステップS103）。画像センサ2が複数ある場合、第2取得部103は各画像センサ2の位置情報203を取得すればよい。
- [0113] 次に、行動選択部102は、位置情報203と間取り情報202とを用いて画像センサ2が設置された空間を特定する（ステップS104）。例えば、行動選択部102は、間取り情報202に含まれる複数の空間のうち、位置情報203が示す座標データがどの空間に位置するかを調べることによって、空間を特定すればよい。また、複数の画像センサ2に対する位置情報203が取得された場合、行動選択部102は、各位置情報203に対応する空間を特定すればよい。
- [0114] 次に、行動選択部102は、行動選択テーブルT1を参照し、特定した空間に対応する対象行動である候補行動を選択する（ステップS105）。ここで、複数の空間が特定された場合は各空間に対応する1又は複数の候補行動が選択される。また、選択された候補行動が複数の場合、各候補行動に対応する認識器111が選択される。
- [0115] 次に、認識器選択部110は、選択した候補行動に対応する認識器111を選択する（ステップS106）。この処理の詳細は図8を用いて後述する。
- [0116] 次に、認識器選択部110は、各画像センサ2の識別子と、各画像センサ2が撮影した画像データ204から認識される候補行動と、候補行動の認識に使用される認識器111とを対応付けたリスト情報205を生成し、メモリ200に格納する（ステップS107）。以上によりリスト情報205が生成される。
- [0117] 図8は、図7のステップS106の処理の詳細を示すフローチャートであ

る。まず、認識器選択部 110 は、ステップ S101 で取得された対象行動のラベルを取得する（ステップ S201）。次に、認識器選択部 110 は、認識器選択テーブル T2 を参照し、対象行動の認識に使用される認識器 111 を選択する（ステップ S202）。ここで、対象行動が複数の場合、対象行動毎に認識に使用される認識器 111 が選択される。次に、認識器選択部 110 は、依存する認識器 111 を選択する（ステップ S203）。ここで、どの認識器 111 がどの認識器 111 に依存するかは、後述する事前知識テーブル T4 の認識器の依存関係の項目を参照することで決定される。

[0118] 図 9 は、行動認識装置 1 における行動認識処理の一例を示すフローチャートである。図 9 は、ある 1 つの画像センサ 2 から 1 フレームの画像データ 204 が撮影されたときの処理が示されている。したがって、画像センサ 2 が複数の場合、図 9 の処理は各画像センサ 2 よって撮影された画像データ 204 に対して並列的に実行される。また、図 9 の処理は、1 フレームの画像データ 204 が撮影される都度、実行されてもよいし、複数フレームの画像データ 204 が撮影される都度、実行されてもよい。

[0119] まず、第 2 取得部 103 は、画像センサ 2 が撮影した画像データ 204 を取得し、メモリ 200 に格納する（ステップ S301）。

[0120] 次に、認識器選択部 110 は、リスト情報 205 を参照し、画像データ 204 を撮影した画像センサ 2 の識別子に対応付けられた候補行動のうち、1 の候補行動を選択する（ステップ S302）。これにより、空間に対応する候補行動が選択されることになる。

[0121] 次に、認識器選択部 110 は、リスト情報 205 を参照し、1 の候補行動の認識に使用される認識器 111 に画像データ 204 を入力し、認識器 111 に特徴量を算出させる（ステップ S303）。算出された特徴量は結合器 121 に入力される。

[0122] 次に、結合器 121 は、1 の候補行動に対応する重み係数で入力された特徴量を重み付ける（ステップ S304）。

[0123] 次に、結合器 121 は、重み付け後の特徴量を連結する（ステップ S30

5)。連結された特徴量は、分類器122に入力される。次に、分類器122は、入力された特徴量の尤度を算出する（ステップS306）。

[0124] 次に、認識器選択部110は、全候補行動が選択済みか否かを判定する（ステップS307）。全候補行動が選択済みでなければ（ステップS307でNO）、処理はステップS302に戻り、次に選択された1の候補行動に対する尤度が算出される。

[0125] 一方、全候補行動が選択済みの場合（ステップS307でYES）、出力部106は、各候補行動に対して算出された尤度のうち、最大の尤度を持つ候補行動のラベルを認識結果として出力する（ステップS308）。なお、ステップS308では、尤度が最大の候補行動のラベルのみならず、尤度の降順にk個の候補行動のラベルを出力してもよい。例えば、kが5であれば、上位5クラスの出力となる。

[0126] このように、本実施の形態によれば、画像センサ2の位置情報203と間取り情報202とに基づいて対象行動の中から認識対象となる候補行動が選択され、候補行動に対する認識結果が算出されている。そのため、空間に依存する人の行動を精度よく認識することが可能となる。

[0127] （認識器の選定）

次に、認識器の選定方法について説明する。従来のDNNによる行動認識方法は、大きく分けて2種類のアプローチがある。

[0128] 第一のアプローチは、汎用的な特徴抽出層である、畳み込み層及びプーリング層を複数用いて、入力データから行動の特徴量を抽出する行動認識方法である。この行動認識方法では、抽出した特徴量から、所与の行動の尤度が算出される。しかしながら、この行動認識方法は、計算コスト及び認識精度との両方において優れたアプローチではない。

[0129] 第二のアプローチは、認識すべき行動に寄与する特徴量をヒューリスティックに設計し、第一のアプローチと同様に、抽出した特徴量から、所与の行動の尤度を算出する行動認識方法である。このヒューリスティックな設計の詳細な説明を以下、説明する。

[0130] 画像処理分野において、人の表現として骨格情報（肩及び膝のような関節間を直線で結んだ表現）が使用され、物体の位置表現として外接矩形（バウンディングボックス）が広く使用されている。このため、前述の骨格情報又は外接矩形を利用したヒューリスティックな行動認識方法が提案されている。しかしながら、骨格情報と外接矩形との表現はともに、行動認識の特徴量として考案されたものではない。したがって、ヒューリスティックな試行錯誤の結果に基づいて、行動認識方法が決定される。また、DNNの学習及び評価に多大な時間が必要なため、試行される条件は限定的である。このようなヒューリスティックな試行錯誤が実施される背景として、認識すべき行動が従来の認識対象と同じか類似しており、従来の認識で使用されていた特徴量を参考にできるためである。特に、公開データセットを学習に使用する場合、認識すべき行動はデータセットで規定されたクラスとなる。

[0131] さらに、従来のDNNによる行動認識方法は、認識すべき行動が従来の認識対象に含まれない場合においても、効果的なアプローチではない。したがって、認識すべき行動に即した特徴の抽出が、効率的な行動認識において必要である。そのためには、認識すべき行動の学習データが与えられた時に、特徴量の抽出方法を決定できる構成が必要となる。

[0132] DNNによる行動認識は、センサデータをDNNで処理することで、所与の行動の尤度を算出する。単一行動のラベルを取得する場合、尤度が最大となる行動のラベルを選択する。複数のDNN又は特徴抽出器から各行動の尤度を算出する場合、特徴量のデータ量が大きい程、認識精度が高くなる保証はない。これは、特徴量の中に、認識精度に貢献せずに、ノイズのように振る舞う成分が含まれる可能性が高くなるからである。同様に、使用するDNN又は特徴抽出器の個数が増加した場合も、必ずしも精度が向上するとは限らない。つまり、DNN又は特徴抽出器の個数を認識すべき行動に適した情報が得られる個数に設定することが重要となる。

[0133] 計算コストを考慮する場合、所望の認識精度を実現する構成の中から、最も計算コストが少ない構成を選択することが重要である。但し、典型的なD

NN又は特徴抽出器は、内部処理に非線形変換を含むため、認識精度を解析的に算出することはできない。換言すれば、ある構成における認識精度を得るためには、データセットを用意し、実際に認識処理を実行する以外の方法はない。このような認識精度を得るための計算コストは大きく、限られた条件数しか試すことはできない。

[0134] 上記で述べたように、計算コストの大きいDNNを用いた行動認識をリアルタイム処理に使用するには、使用するDNN又は特徴抽出器の構成が重要となる。

[0135] 畳み込みの処理をするDNNの層（畳み込み層）の計算コストを以下に説明する。数学における畳み込み処理と同様に、畳み込み層は所与のカーネルを入力画像データに対し適用することで出力を得る。畳み込み層は1フレームの二次元の画像データに対して適用する二次元畳み込みと、Nフレーム（Nはフレーム数）の二次元の画像データに対して適用する三次元畳み込みが存在する。畳み込みの重み係数のパラメータ数 W_c は、カーネルサイズを k 、畳み込み次元数を d （二次元畳み込みの場合は2、三次元畳み込みの場合は3）、入力のチャンネル数を C_i 、出力のチャンネル数を C_o とすると、式（3）で表せる。

[0136] [数3]

$$W_c = k^d \times C_i \times C_o \quad (3)$$

畳み込み計算に必要な乗算数 U_c は、カーネルの移動量であるストライド数を s とすると、式（4）で表せる。説明を簡単にするため、画像の縦方向と横方向の処理は同一とする。

[0137] [数4]

$$U_c = W_c \times (C_i/s)^d \quad (4)$$

一方、全結合層の重み係数の個数 W_{fc} は、式（5）で示すように入力チ

チャンネル数 C_i と出力チャンネル数 C_o との積になる。

[0138] [数5]

$$W_{fc} = C_i \times C_o \quad (5)$$

全結合層の計算に必要な乗算数 U_{fc} は、式 (6) に示すように重み係数の個数 W_{fc} と等しい。

[0139] [数6]

$$U_{fc} = W_{fc} \quad (6)$$

重み係数の個数 W_{fc} は実行に必要なメモリ数に相当し、乗算数は実行に必要な演算数に相当する。式 (5) に示すように、畳み込みの演算数は、カーネルに起因する乗算により全結合より大きくなる。3次元畳み込みの場合、重み係数の個数、乗算数がともに増大する。したがって、図2で示すような多層からなる畳み込み層を中心としたDNNの計算コストは大きくなり、デスクトップ型の計算機にグラフィックスプロセッサユニット (GPU) を用いた場合においても、リアルタイム処理とみなせる24fps (一般的な映画相当) の処理が困難な場合もある。認識精度向上のために、DNNを複数使用した場合、リアルタイム処理はさらに困難になる。

[0140] 従来のDNNの計算コストが大きい理由は、ネットワークを構成する層を主に直列に並べ、その大半の層が畳み込み層からなる点にある。ネットワークの中間層においても、重み係数の個数が多く、乗算数も増加している。中間層の出力は人間による解釈が困難なテンソルとなっており、中間層の出力が必ずしも認識に最適な表現となっていない。換言すると、中間層に必要以上の重み係数が存在し、乗算数も増加している。すなわち、中間層において、人の行動或いは空間の状態を少ないパラメータで効果的に表現することにより、重み係数の個数の減少及び乗算数の減少を図ることが可能となる。

[0141] 人の行動を識別する場合を考えると、密なデータである入力画像ではなく

、人の骨格情報からも行動を推定することは可能である。例えば、包丁で食材を切るという行動は、特徴的な手の上下移動の反復である。入力画像データの表現に必要なデータ数（バイト数）は、8 b i t の画像データの場合、画像データの縦の長さ×横の長さ×チャンネル数との積となる。一方、二次元の骨格情報であれば、骨格の頂点（例えば、17点）の個数と、骨格の頂点の座標データ（ x , y ）の成分数との積である34で表現可能となる。

[0142] 上記を考慮し、以下、本発明の実施の形態における認識器111の選定方法を説明する。認識対象となる人の行動は例えば、歩く、走る、立つ、会話する、掃除する、料理を開始するといった動作の他、例えば、寝ている、横になっている、座っている、テレビを見ている等の静止した行動も含まれる。

[0143] まず、認識器111の選択肢が多いことについて説明する。第1行動認識部104で使用する認識器111の個数は少なくとも一つである。第1行動認識部104がN個の認識器111を保持し、M個の認識器111を選択する場合の認識器111の組み合わせはNの階乗オーダーとなる。Nの階乗オーダーは指数関数よりも発散が速い。学習時間を考慮すると、Nが5程度より大きくなると全ての認識器111の組み合わせを学習することが困難になる。

[0144] 次に、認識器111の組み合わせに関する問題を、数学的な問題として表現する。選択された認識器111の組み合わせ毎に、学習を行うことで認識精度と計算コストとが得られる。満たすべき認識精度（例えば、70%）が与えられ、その認識精度を満たす中で、計算コストが最小となる組み合わせを探すことは、条件付きの組み合わせ最適化問題となる。このような組み合わせ最適化問題は、組み合わせの個数が全探索の困難な個数になると、解析的に解くことができなくなる。したがって、準最適解を探索することになる。なお、組み合わせ最適化問題に対する万能なアルゴリズムは存在しない。

[0145] そこで、認識器111の選択を効率的に探索するために、事前知識と貪欲法とに基づき認識器111を選択する。

- [0146] まず、事前知識に関して説明する。図10は、認識器111に関する情報を纏めた事前知識テーブルT4のデータ構成を示す図である。このテーブルはメモリ200に記憶されている。事前知識テーブルT4は、認識器111の識別番号、認識内容、入力センサデータ、相対計算コスト、及び認識器の依存関係に関する項目を有する。各項目について説明する。識別番号は登録されている認識器111に対して重複のない番号（例えば、連番）が与えられ、組み合わせ探索におけるインデックスまたはハッシュとなる。認識内容は、認識器111が認識する内容である。入力センサデータは認識器111が推論をするために必要となるセンサデータであり、0個以上である。0を含むのは、依存する他の認識器111が算出した特徴量のみで推論する認識器111は、センサデータを必要としないためである。相対計算コストは、認識器111の計算コストの相対値である。相対計算コストは、基準となる計算機で実行されたベンチマーク結果を元に算出された相対値である。なお、相対計算コストは、複数の計算機のベンチマーク結果に基づいて算出されてもよいし、ベンチマーク結果そのものである絶対値が採用されてもよい。
- [0147] 認識器の依存関係は、他の認識器111に依存するか否かを示す情報である。例えば、識別番号が4～6の認識器111は識別番号が3の認識器111に依存する。これは、頭向き推定、年齢性別推定、及び個人推定を行う場合、顔検出の結果が利用されるからである。
- [0148] なお、手又は足の検出する認識器111のように、人の特定領域を検出する認識器111が用いられてもよい。また、画像の範囲を示すマスクのような密なデータを出力する認識器111が用いられてもよい。
- [0149] 次に、貪欲法について説明する。貪欲法とは、部分的な情報に基づき、最適な選択をする手法である。認識器111の選定では、まず、事前知識テーブルT4に含まれる各認識器111について、認識器111を一つのみ選択した時の認識精度を、学習データを用いて計算する。ある一つの認識器111のみ使用した場合の認識精度が高い場合、その認識器111を他の認識器111と組み合わせて使用した場合の認識精度も高いことが期待される。但

し、その保証はない。貪欲法では、認識器 1 1 1 を一つのみ使用した場合の認識精度が高い場合、その認識器 1 1 1 を他の認識器 1 1 1 と組み合わせて使用した場合の認識精度も高いと仮定し、評価対象となる認識器 1 1 1 の組み合わせの優先順位が決定される。具体的には、認識器 1 1 1 を一つのみ使用した場合の認識精度の降順に認識器 1 1 1 をソートし、先頭の認識器 1 1 1 から優先的に評価する。例えば、先頭を 1 番目の認識器 1 1 1 とし、順にインデックスを振った場合、1 と 2、1 と 3、1 ～ 3、1 ～ 4、のように、先頭に近いインデックスを優先的に使用する。より詳細なインデックスの選択方法は、先頭から X 番目のインデックスまで考慮するパラメータ X と、Y 個までのインデックスの選択を許容するパラメータ Y で表現できる。この X と Y のパラメータは推論を実行する計算機の性能に応じて決定され、一例として、X = 4、Y = 3 がある。図 5 に示す認識器選択テーブル T 2 は、例えば以上のような認識器 1 1 1 の選定結果に基づいて作成される。

[0150] (分類器の選定)

次に、分類器 1 2 2 の選定について説明する。第 1 行動認識部 1 0 4 が出力する特徴量は骨格情報のように、少ないパラメータをベクトル化したデータであるため、CNN を用いなくても識別可能である。分類器 1 2 2 は CNN に比べて計算処理量が大幅に少ない。複数の分類器の候補に対して交差検証（クロスバリデーション）を行い、正確度（Accuracy）又は F1 スコアが最大となる分類器 1 2 2 を選定すればよい。候補となる分類器 1 2 2 は、例えばロジスティクス回帰、サポートベクターマシン（SVM）、決定木、ランダムフォレスト、k 近傍法、ガウシアンナイーブベイズ、パーセプトロン、又は確率的降下法を用いた分類器である。

[0151] (分類器の学習)

次に、分類器 1 2 2 の学習について説明する。分類器 1 2 2 は対象行動のラベルを含む学習データを用いて学習すれば良い。例えば、「料理を開始する」と「料理を開始していない」とのそれぞれのラベルに対して学習データとなる動画の画像データを用意する。用意した画像データを用いて、図 9 に

示す行動認識処理を行動認識装置 1 に実行させ、得られたラベルと学習データの正解ラベルとを比較し、その誤差が最小になるように、分類器 122 の重みを更新する。

[0152] (シミュレーション)

次に、行動認識装置 1 が、既存の行動推定 A1 より計算処理量が少ないことを確認するために行ったシミュレーションについて説明する。

[0153] このシミュレーションでは、既存の認識器 111 として CNN が用いられた。具体的には、姿勢推定の認識器 111 として「1: PoseNet」を用い、物体検出の認識器 111 として、「2: SSD」を用い、顔検出の認識器 111 として「3: RetinaFace」を用い、頭向き推定の認識器 111 として「4: DeepHeadPose」を用いた。分類器 122 としては確率的降下法の分類器が用いられた。一方、既存の行動推定 A1 として「5: RepresentationFlowNet」が用いられた。

[0154] 入力画像データとしては、3 名の人を含む 1 枚のカラーの画像データが用いられた。この画像データは、解像度が 640×480 、チャンネル数が 3、分解能が 8 bit であった。分類器 122 の計算コストは、識別するクラス数によらず、既存の認識器 111 に比べて十分小さいため無視した。

[0155] 計算機は、グラフィックスプロセスユニット（製品型番: GeForce GTX 1080Ti）による計算補助を含むデスクトップ型のパーソナルコンピュータが使用された。

[0156] 図 11 は、既存の認識器 111 を個別に実行した場合の 1 フレームあたりの処理時間を纏めたテーブルである。シミュレーションでは、図 11 の処理 A（姿勢推定）、処理 B（物体検出）、処理 C（頭向き検出）のそれぞれの認識器 111 を実行したときの処理時間が計測された。処理 C の処理時間は頭向き推定と顔検出との処理時間が含まれる。処理 A、B、C は互いに独立した処理のため、並列実行可能である。オーバーヘッドを無視した場合、処理 A、B、C のうち処理時間の最大値は処理 C の 0.0455 秒であった。一方、処理 A、B、C を逐次処理した場合の全体の処理時間は 0.0725

秒であった。

[0157] これに対し、処理D（既存の行動推定A1）の処理時間は0.1429秒であった。したがって、処理A、B、Cを並列処理した場合の処理速度は、処理Dの3倍であった。処理A、B、Cを逐次処理した場合の処理速度は処理Dの2倍であった。このように、複数の認識器111を用いた方が、既存の行動推定A1よりも処理速度が大幅に速くなることが確認できた。

[0158] 次に、行動認識装置1の認識精度を評価するために行われたシミュレーションについて説明する。このシミュレーションでは、入力画像データとして図12に示す5本の動画の画像データが用いられた。図12は、入力画像データを纏めた表である。ビデオIDは5本の動画を識別するための識別子である。図12では、各動画について「料理を開始する」のフレームと「料理を開始していない」のフレームとのそれぞれの総数が示されている。

[0159] 図13は、行動認識装置1の認識精度を評価するために行われたシミュレーションの結果を纏めた表である。図13において、例えば処理A+B+Cは、第1行動認識部104を処理Aの認識器111、処理Bの認識器111、処理Cの認識器111で構成した場合を示し、例えば処理A+Bは、第1行動認識部104を処理Aの認識器111と処理Bの認識器111とで構成した場合を示している。このシミュレーションでは、図13で示す処理A～Cの組み合わせのそれぞれについて、8種類の分類器122のそれぞれの認識精度及び処理時間が計測された。8種類の分類器122は、ロジスティクス回帰、サポートベクターマシン、決定木、ランダムフォレスト、k近傍法、ガウシアンナイーブベイズ、パーセプトロン、及び確率的降下法である。このシミュレーションでは、8種の分類器122のうち、確率的降下法の分類器122を採用した場合の処理時間が最も短く、認識精度も最大であった。

[0160] また、確率的降下法の分類器122を用いた場合において、図13に示す処理A～Cの各組み合わせのうち認識精度の平均値は88.5%であった。この結果は、行動認識装置1の認識精度が「料理を開始する」と「料理を開始

していない」との２クラス分類をランダムに推定した場合の認識精度の期待値である５０％より高いことを示す。したがって、このシミュレーション結果から行動認識装置１は、行動を認識できることが確認できた。

[0161] 図１３に示す各組み合わせのうち、認識精度が最良なのは、処理Ａ＋Ｃで８８．８％であった。しかし、処理Ａのみでも８８．７％の認識精度が得られた。仮に認識精度の閾値を８８．７％とすると、処理Ａのみ用いた場合、計算コストが最小になる。処理Ａ＋Ｂ＋Ｃが最良ではないのは、精度の向上に貢献しない認識器１１１があり、その認識器１１１がノイズ源のように振る舞ったからである。

[0162] 以上のシミュレーション結果は、一例であり、第１行動認識部１０４を構成する各認識器１１１は上述したものに限定されず、対象行動に応じて適宜最適な認識器１１１が採用される。また、分類器１２２も確率的降下法以外の分類器１２２が採用されてもよい。

[0163] （変形例）

[0164] （１）上記実施の形態では、行動認識装置１は、単体で構成されたが、本開示はこれに限定されず、複数の装置で構成されてもよい。図１４は、本開示の変形例に係る行動認識装置１Ａの構成の一例を示すブロック図である。行動認識装置１Ａは、サーバで構成されている。行動認識装置１Ａは、通信部５００をさらに備える。通信部５００は、出力部１０６が出力する認識結果をネットワークＮＴ及びゲートウェイ７００を介して家電機器６００に送信する。通信部５００は、宅内に設置された画像センサ３００が撮影した画像データを受信する。ネットワークＮＴは、例えばインターネット等の広域通信網である。ゲートウェイ７００は宅内に設置され、画像センサ３００及び家電機器６００をネットワークＮＴに接続する。家電機器６００は、洗濯機、電子レンジ、及びテレビ等である。家電機器６００は、行動認識装置１Ａから送信された行動の認識結果を利用した制御を実行したり、認識結果を表示したりする。このように、行動認識装置１Ａはサーバで構成された場合においても、宅内の人の行動を認識できる。

[0165] (2) 上記実施の形態では、入力データとして画像データが用いられたがこれに加えて、熱画像データ、奥行画像データ、音声データ、室温データ、湿度データ、照度データ、無線の電波のデータのうちの少なくとも1つが用いられてもよい。

[0166] (3) 宅内における人の行動を宅内に設置したカメラで認識する場合、カメラの設置位置及びアングルは、固定又は低い頻度で変更されることが想定される。また、カメラの視野角は広角レンズの場合で110°程度、狭角レンズの場合はそれ以下となり、あるカメラに映る範囲は設置された空間の一部となる。このため、認識したい行動の発生頻度が高い、或いは発生時の重要度が高い空間が撮影されるようにカメラは設置されることが想定される。これらの想定において、カメラが設置されている空間と、カメラに映る或いは映る可能性がある物体とを事前に知ることができれば、行動の認識精度が向上する。例えば、発生頻度が0の行動は、認識対象から除外することができる。また、移動頻度が低いソファがカメラに映っている場合はそこに座るという行動が発生する頻度が高いことが想定される。また、移動可能な椅子が空間内に存在している場合、椅子に座るという行動の頻度が高いことが想定される。これにより、図6に示す重みテーブルT3の重み係数の値を決定することが可能となる。

[0167] 宅内における行動認識において、人と物体との直接的な接触或いは物体の近傍に人がいるという人と物体とのインタラクションが重要である。物体の移動、開閉、操作を伴う物体（後述の第1物体）の設置位置が既知の場合、人がその物体に対してどのような位置関係又は向きにいるかということが、行動認識の判断材料となる。これは人が行動の際に、物体の移動、開閉、及び操作を行うことが多いことに基づく。

[0168] 水回り設備に代表される宅内の移動できない設備（後述の第2物体）も行動認識において重要である。料理又は洗面といった水を使用する行動は通常、宅内に設置された設備を使用する。そのため、水回りの設備の位置と、人の位置関係とが行動認識の判断材料となる。これは人の生活において、水を

必要とする行動の頻度が高いということに基づく。

[0169] 住居は一般的に4～10個程度の部屋にドア又は固定若しくは可動式の壁で区切られる。また、多層式の住居は、階段等の昇降設備を備える。このため、住居は複数の空間及び多数のドアを含む物体（後述の第3物体）を有する。これらの空間の名称と、ドア又は昇降設備の位置と、人との位置関係が行動認識の判断材料となる。これは、調理又は入浴のように特定の部屋でのみ実施する行動があることに基づく。

[0170] 上述したように、物体又は設備がどこにあり、どのように空間が区切られているかという情報が行動認識において重要である。従来、リビングや寝室といった空間名を考慮に入れて行動認識を行う技術はあったが、同じ空間名であっても使われ方は多様であり、リビング又は寝室という空間名のみでは、行動認識の判断材料としては不十分である。そのため、どの空間でどのような人と物体とのインタラクションが起こり得るかを考慮することが重要である。これらのことを踏まえて、以下、本開示の変形例（3）について説明する。

[0171] 図15は、本開示の変形例（3）に係る行動認識装置1Bの構成の一例を示すブロック図である。この変形例は、間取り情報202に基づいて、間取り特徴量を抽出し、抽出した間取り特徴量に基づいて行動選択テーブルT1を生成するものである。この変形例では間取り情報202は、建物内の空間の情報に加えて、各空間に設置された物体（設備及び機器）の情報（種別情報及び位置情報）も含んでいる。

[0172] 行動認識装置1Bは、表示端末400と所定の通信路を介して通信可能に接続されている。所定の通信路としては、無線LAN、Bluetooth（登録商標）が採用できる。また、所定の通信路は、インターネットであってもよい。表示端末400は、例えば、スマートフォン又はタブレット型コンピュータで構成される。表示端末400は、例えばユーザにより所持される。ユーザは例えば、画像センサ2の施工者である。施工者は、例えば住居の居住者又は施工業者である。

- [0173] 行動認識装置 1 B のプロセッサ 1 0 0 B は、図 1 に対してさらに、設置支援部 3 0 1 及び間取り特徴量抽出部 3 0 2 を含む。間取り特徴量抽出部 3 0 2 は、間取り情報 2 0 2 に基づいて、建物に設置された物体を抽出し、抽出した物体を、移動可能な第 1 物体、水回り設備である第 2 物体、建物の構造物である第 3 物体のいずれかに分類し、分類した物体のそれぞれについて、設置位置と、分類結果を示す分類情報とが対応付けられた間取り特徴量を抽出する。間取り特徴量は、例えば 2 次元テーブルで構成される。
- [0174] 第 1 物体は、例えば、家具及び電気製品等の移動可能な物体を含む。具体的には、第 1 物体は、掃除機、コーヒーマーカ、ノートパソコン、椅子、及びソファなどである。
- [0175] 第 2 物体は、例えばシンク及び洗面台等の移動不可能な水回り設備などである。
- [0176] 第 3 物体は、玄関、キッチン、リビング、ダイニング、ベッドルーム、バスルーム、昇降設備、及び洗面所等の建物の構造物である。
- [0177] 設置位置は、例えば建物の玄関を基準とする 2 次元又は 3 次元の座標データで構成される。なお、第 3 物体の設置位置は、第 3 物体が位置する領域を示す座標データで構成されてもよい。これにより、第 1 物体及び第 2 物体が建物内のどの空間に設置するかを判定できる。なお、第 1 物体及び第 2 物体の設置位置は、第 1 物体及び第 2 物体が位置する座標データに加えて、第 1 物体及び第 2 物体が位置する建物の空間の名称を含んでもよい。
- [0178] 分類情報は、間取り情報 2 0 2 から抽出される物体が、第 1 物体～第 3 物体のいずれに該当するかを示す情報である。
- [0179] 間取り特徴量抽出部 3 0 2 は、抽出した間取り特徴量に基づいて、行動選択テーブル T 1 を生成する。
- [0180] 例えば、第 1 物体の設置位置は、ユーザが動作する位置を示す。例えば、コーヒーマーカの設置位置はコーヒを淹れる行動に対応付けることができる。例えば、電子レンジの設置位置は、ユーザが料理を開始する行動に対応付けることができる。

- [0181] また、第2物体は水回り設備であるため、第2物体の設置位置は、水を使用する動作が行われる位置を示す。例えば、シンクの設置位置は、皿を洗う行動に対応付けることができる。
- [0182] 第3物体の設置位置は、空間の名称と、ユーザの空間への出入りと、ユーザが移動可能な範囲を示す。例えば、浴室のドアの設置位置は入浴動作に対応付けることができる。
- [0183] 間取り特徴量抽出部302は、抽出した間取り特徴量に基づいて、行動選択テーブルを生成する。行動選択テーブルT1は、図4に示すように建物内の1以上の空間と各空間においてユーザが行う可能性の高い対象行動とが対応付けられたテーブルである。
- [0184] ここで、間取り特徴量抽出部302は、間取り特徴量から、各空間にどのような機器又は設備が設置されているかを決定し、決定結果と分類情報とから予め定められたルールを参照することで、行動選択テーブルT1を生成すればよい。ルールとしては、例えば、第1物体である電子レンジに対して料理を温める、第1物体であるコーヒーメーカーに対してコーヒーを淹れる、第2物体であるシンクに対して皿洗いというように、機器又は設備と機器又は設備で行われる可能性の高い対象行動とが対応付けられたルールが採用できる。
- [0185] 例えば、キッチンに、電子レンジ、コーヒーメーカー、及びシンクが設置されているとする。この場合、間取り特徴量抽出部302は、キッチンについて、料理を開始する、コーヒーを淹れる、皿洗いをするといった対象行動が対応付けられた行動選択テーブルT1を生成すればよい。なお、間取り特徴量抽出部302は、建物内の他の空間についても同様にして対象行動を対応付ければよい。
- [0186] また、ルールは、機器又は設備のみならず、空間自体に対して行われる可能性の高い対象行動を含んでいてもよい。例えば、図4で説明したように、キッチンに対して料理を開始する、洗面所に対して洗濯、リビング、ダイニング、及びベッドルームに対して読書、リビング及びダイニングに対して食

事といったルールが採用できる。

[0187] さらに、間取り特徴量抽出部 302 は、抽出した間取り特徴量から図 6 に示す重みテーブル T3 に記載された重み係数を設定してもよい。例えば、キッチンに、シンク、冷蔵庫、電子レンジ、及びガスコンロが設置されている場合、キッチンについての重み係数 W_e は、シンク、冷蔵庫、電子レンジ、及びガスコンロについて 1、それ以外の機器については 0 に設定すればよい。

[0188] 間取り特徴量抽出部 302 は、キッチン以外の他の空間についてもキッチンと同様に重み係数 W_e を設定すればよい。

[0189] なお、第 1 取得部 101 は、このようにして生成された行動選択テーブル T1 を対象行動情報として取得すればよい。

[0190] 設置支援部 301 は、画像センサ 2 が設置される建物の空間の名称を表示端末 400 を介して取得し、空間に関連する特定機器又は特定設備が画像センサ 2 の視野内に含まれるように画像センサ 2 を設置する設置ガイダンスを表示端末 400 に出力する設置支援部をさらに備える。

[0191] 図 16 は、玄関 501 に画像センサ 2 が設置されるシーンを示す図である。図 17 は、玄関 501 に画像センサ 2 が設置されるシーンにおいてユーザと表示端末 400 とのインタラクションの一例を示す図である。

[0192] ステップ S T1 において、表示端末 400 は、設定画面を開く操作をユーザから受け付け、画像センサ 2 の設置が予定される空間を一覧表示する。ここでは、A：玄関、B：キッチン、C：リビング、・・・というように画像センサ 2 が設置される空間の名称が一覧表示される。

[0193] ステップ S T2 において、表示端末 400 は、一覧表示された空間の名称の中から画像センサ 2 が設置される空間の名称を選択するユーザの操作を受け付ける。表示端末 400 は、選択された空間の名称を行動認識装置 1B に送信する。ここでは、画像センサ 2 は玄関 501 に設置されるため、「A：玄関」が選択される。

[0194] ステップ S T3 において、表示端末 400 は、「ドアが映る位置に設置し

て下さい。設置後、OKを押して下さい。」というように、ユーザに画像センサ2の設置を促すメッセージを表示する。設定画面には、このメッセージに併せて、ユーザが画像センサ2の設置作業が完了したことを表示端末400に知らせるためのOKボタンが表示されている。

[0195] この設置作業により、図16に示すように、玄関501に画像センサ2が設置される。

[0196] ステップST4において、表示端末400は、OKボタンを押すユーザの操作を受け付ける。OKボタンを押す操作を受け付けた表示端末400は、OKボタンが押されたことを示す情報を行動認識装置1Bに送信する。

[0197] OKボタンが押されたことを示す情報を取得した設置支援部301は、画像センサ2が撮影した画像データを取得し、取得した画像データからドア401を検出する処理を実行する。このとき、設置支援部301は、表示端末400に、「ドアを自動検出しています。」というメッセージを表示させる（ステップST5）。ここで、設置支援部301は、所定の認識器を用いてドア401を検出すればよい。ドア401は玄関501における特定設備の一例である。特定設備は、画像センサ2が設置される空間に応じて予め定められている。所定の認識器としては、例えばドア401を検出するために予め準備された認識器が採用される。

[0198] 図18は、図17に続くインタラクションを示す図である。ステップST6において、表示端末400は、ドア401の検出結果を示すアノテーション情報を設置支援部301から取得し、取得したアノテーション情報が示すアノテーション画像を、画像センサ2が撮影した画像に重畳表示する設定画面を表示する。この設定画面には、「自動検出された赤色ドアガイドの位置のドアに合わせるように調整し、OKを押してください」というように、アノテーション画像の修正を促すメッセージが表示されている。アノテーション情報は、例えばアノテーション画像の座標データである。

[0199] 図19は、アノテーション画像A1が重畳された設定画面G1、G2の一例を示す図である。設定画面G1は修正前のアノテーション画像A1を示し

、設定画面 G 2 は修正後のアノテーション画像 A 1 を表示する。アノテーション画像 A 1 は、四角形のバウンディングボックスで構成されている。アノテーション画像 A 1 は所定の色（ここでは、赤色）を有している。設定画面 G 1、G 2 には、アノテーション画像 A 1 の修正作業の完了を告げる OK ボタン 1901 が表示されている。

[0200] アノテーション画像 A 1 の左上の頂点には丸マーク P 1 が表示され、アノテーション画像 A 1 の右下の頂点には丸マーク P 2 が表示されている。

[0201] 表示端末 400 は、丸マーク P 1、P 2 を移動する操作を受け付けると、受け付けた操作に連動させてアノテーション画像 A 1 のサイズを変更する。

[0202] 設定画面 G 1 では認識器によるドア 401 の検出結果であるアノテーション画像 A 1 が表示されている。このアノテーション画像 A 1 は、ドア 401 よりも左斜め上側にずれており、ドア 401 が正しく認識できていないことが分かる。

[0203] ユーザは、アノテーション画像 A 1 の形状がドア 401 の外接矩形となるように、丸マーク P 1、P 2 を移動する操作を入力する。これにより、設定画面 G 2 が得られる。設定画面 G 2 では、アノテーション画像 A 1 がドア 401 の外接矩形に位置決めされている。

[0204] これにより、設定画面 G 1 においてドア 401 からずれた位置に表示されていたアノテーション画像 A 1 が、設定画面 G 2 に示すようにドア 401 の全域を取り囲むように修正される。その結果、認識器がドア 401 の画像上の位置を正確に認識できていない場合においても、認識器にドア 401 の画像上の正しい位置が得られる。

[0205] ステップ S T 7 において、表示端末 400 は、アノテーション画像 A 1 の調整が完了したユーザから、OK ボタン 1901 を押す操作を受け付ける。この操作を受け付けた表示端末 400 は、修正されたアノテーション画像 A 1 の座標データを設置支援部 301 に送信する。この座標データを受信した設置支援部 301 は、修正されたアノテーション画像 A 1 の座標データを玄関 501 に設置された画像センサ 2 の位置情報 203 と対応付けてメモリ 2

００に記憶する。

- [0206] 以後、玄関５０１に対応する対象行動を認識するための認識器１１１は、玄関５０１に設置された画像センサ２により撮影された画像データに対して、座標データが示すアノテーション画像Ａ１を基準にしてユーザの行動を認識する。これにより、認識器１１１は効率良くユーザの行動を認識することができる。
- [0207] ステップＳＴ８において、表示端末４００は、設置支援部３０１から座標データを保存したことを示す情報を受信したため、「設定を保存しました。」とのメッセージを表示する。
- [0208] ステップＳＴ９において、表示端末４００は、ユーザから設定画面を閉じる操作を受け付ける。これにより、表示端末４００は、設定画面を閉じる。
- [0209] 図２０は、キッチン５０２に画像センサ２が設置されるシーンを示す図である。図２１は、キッチン５０２に画像センサ２が設置されるシーンにおいてユーザと表示端末４００とのインタラクションの一例を示す図である。
- [0210] ステップＳＴ１は、図１７のステップＳＴ１と同じである。ステップＳＴ２において、表示端末４００は、ユーザから画像センサ２の設置場所として「Ｂ：キッチン」を選択する操作を受け付ける。
- [0211] ステップＳＴ３は、図１７のステップＳＴ３と同じである。ステップＳＴ４において、表示端末４００は、ＯＫボタンを押すユーザの操作を受け付ける。ＯＫボタンを押す操作を受け付けた表示端末４００は、ＯＫボタンが押されたことを示す情報を行動認識装置１Ｂに送信する。
- [0212] ＯＫボタンが押されたことを示す情報を取得した設置支援部３０１は、画像センサ２が撮影した画像データを取得し、取得した画像データから冷蔵庫４０２を検出する処理を実行する。これにより、設置支援部３０１は、表示端末４００に、「冷蔵庫を自動検出しています。」というメッセージを表示させる（ステップＳＴ５）。ここで、設置支援部３０１は、所定の認識器を用いて冷蔵庫４０２を検出すればよい。冷蔵庫４０２はキッチン５０２における特定機器の一例である。特定機器は、画像センサ２が設置される空間に

応じて予め定められている。所定の認識器としては、例えば冷蔵庫４０２を検出するために予め準備された認識器が採用される。

[0213] 図２２は、図２１に続くインタラクションを示す図である。ステップＳＴ６～ＳＴ９は図１８と同じである。図２３は、アノテーション画像Ａ１が重畳された設定画面Ｇ３、Ｇ４の一例を示す図である。設定画面Ｇ３は修正前のアノテーション画像Ａ１を表示し、設定画面Ｇ４は修正後のアノテーション画像Ａ１を表示する。設定画面Ｇ３では認識器による冷蔵庫４０２の検出結果であるアノテーション画像Ａ１が表示されている。このアノテーション画像Ａ１は、冷蔵庫４０２よりも上側にずれており、冷蔵庫４０２が正しく認識できていないことが分かる。ユーザは、アノテーション画像Ａ１の形状が冷蔵庫４０２の外接矩形となるように、丸マークＰ１、Ｐ２を移動する操作を入力する。これにより、設定画面Ｇ４が得られる。設定画面Ｇ４では、アノテーション画像Ａ１が冷蔵庫４０２の外接矩形に位置決めされている。

[0214] これにより、設定画面Ｇ３において冷蔵庫４０２からずれた位置に表示されていたアノテーション画像Ａ１が、設定画面Ｇ４に示すように冷蔵庫４０２の全域を取り囲むように修正される。その結果、認識器が冷蔵庫４０２の画像上の位置を正確に認識できていない場合においても、冷蔵庫４０２の画像上の正しい位置が得られる。

[0215] 以後、冷蔵庫４０２に対応する対象行動を認識するための認識器１１１は、冷蔵庫４０２に設置された画像センサ２により撮影された画像データに対して、座標データが示すアノテーション画像Ａ１を基準にしてユーザの行動を認識する。これにより、認識器１１１は効率良くユーザの行動を認識することができる。

[0216] 図２４は、画像センサ２の設置後にアノテーション画像Ａ１を修正する際のユーザと表示端末４００とのインタラクションの一例を示す図である。画像センサ２は設置後に何らかの外力が加えられてアングルが変動することがある。この場合、設置時に設定されたアノテーション画像Ａ１が示す座標データは、特定機器又は特定設備の画像上の位置からずれてしまう。このずれ

を修正するために以下の修正作業が行われる。

- [0217] ステップS T 1において、表示端末4 0 0は、設定画面を開く操作をユーザから受け付け、画像センサ2の位置ずれの確認を促すメッセージを表示する。ここでは、「カメラ位置がずれていないか確認してください。赤色ガイドが現在の設定です。水色ガイドが自動検出した位置です」とのメッセージが表示されている。
- [0218] 図2 5は、アノテーション画像A 1が重畳表示された設定画面G 5、G 6の一例を示す図である。設定画面G 5は、修正前のアノテーション画像A 1を示し、設定画面G 6は、修正後のアノテーション画像A 1を示している。アノテーション画像A 1は赤色ガイドに相当する。点線のアノテーション画像A 2は認識器による冷蔵庫4 0 2の検出結果であるアノテーション画像を示している。アノテーション画像A 2は水色ガイドに相当する。
- [0219] 設定画面G 5において、画像センサ2の角度が設置時の角度に対してずれたため、アノテーション画像A 1は、アノテーション画像A 2に対して上側にずれて表示されている。
- [0220] ステップS T 2において、ユーザは設定画面G 5から画像センサ2の角度のずれを確認し、画像センサ2の角度の調整作業を行う。
- [0221] 表示端末4 0 0は、アノテーション画像A 1の位置がアノテーション画像A 2の位置と一致するか否かを画像処理により判定する。表示端末4 0 0は、一致したと判定した場合、画像センサ2の角度が設置時の角度に戻ったことを示すメッセージを表示する。
- [0222] ステップS T 4において、表示端末4 0 0は、ユーザから設定画面を閉じる操作を受け付ける。これにより、表示端末4 0 0は、設定画面を閉じる。
- [0223] このような調整作業をユーザに行わせることで、設置後に画像センサ2の角度が変動した場合であっても、画像センサ2の角度を設置時の角度に戻すことができる。これにより、認識器1 1 1はユーザの行動を正確に認識することができる。

産業上の利用可能性

[0224] 本発明の行動認識装置は、建物において人の行動を認識する際に有用である。

請求の範囲

- [請求項1] 建物における人の行動を認識する行動認識装置であって、
予め定められた認識対象となる1以上の対象行動を含む対象行動情報と、前記建物の間取り情報と、前記建物に設置された画像センサの位置情報とを取得する第1取得部と、
前記画像センサの位置情報と前記間取り情報とに基づいて前記対象行動情報に含まれる前記1以上の対象行動の中から認識候補となる候補行動を選択する行動選択部と、
前記画像センサが検出した画像データを取得する第2取得部と、
前記候補行動に応じた1以上の認識器を決定し、前記1以上の認識器を用いて前記画像データの特徴量を算出する第1行動認識部と、
前記特徴量に基づいて前記候補行動を認識する第2行動認識部と、
前記第2行動認識部による認識結果を出力する出力部とを備える、
行動認識装置。
- [請求項2] 前記第1行動認識部は、前記候補行動が所定の行動である場合、複数の認識器を決定し、
前記第2行動認識部は、前記複数の認識器がそれぞれ算出した特徴量を結合し、結合した特徴量に基づいて前記候補行動を認識する、
請求項1記載の行動認識装置。
- [請求項3] 前記所定の行動は、掃除、歯磨き、料理、洗濯、コンピュータの使用、読書、又は食事である、
請求項2記載の行動認識装置。
- [請求項4] 各認識器は、コンボリューションニューラルネットワークで構成され、
前記第2行動認識部は、ロジスティクス回帰、サポートベクターマシン、決定木、ランダムフォレスト、k近傍法、ガウシアンナイーブベイズ、パーセプトロン、及び確率的降下法のいずれか1つを用いた分類器を用いて前記候補行動を認識する、

請求項 1 ～ 3 のいずれかに記載の行動認識装置。

[請求項5] 前記第 2 行動認識部は、前記特徴量を説明変数とし、前記対象行動を目的変数として機械学習された分類器を用いて前記候補行動を認識する、

請求項 1 ～ 4 のいずれかに記載の行動認識装置。

[請求項6] 前記第 2 行動認識部は、前記候補行動に応じて予め定められた重み係数を用いて各特徴量に重み付けをし、重み付け後の各特徴量に基づいて前記候補行動を認識する、

請求項 1 ～ 5 のいずれかに記載の行動認識装置。

[請求項7] 前記間取り情報から、前記建物に設置された 1 以上の物体を抽出し、前記 1 以上の物体を、移動可能な第 1 物体、水回り設備である第 2 物体、及び前記建物の構造物である第 3 物体のいずれかに分類し、前記 1 以上の物体のそれぞれについて、分類結果を示す分類情報と設置位置とが対応付けられた間取り特徴量を抽出し、前記間取り特徴量に基づいて、行動選択テーブルを生成し、

前記行動選択テーブルは、前記建物の 1 以上の空間と、各空間に対応する前記対象行動とが対応付けられたテーブルであり、

前記第 1 取得部は、前記行動選択テーブルを前記対象行動情報として取得する、

請求項 1 ～ 6 のいずれかに記載の行動認識装置。

[請求項8] 前記行動認識装置は表示端末と通信可能に接続され、

前記画像センサが設置される前記建物の空間の名称を前記表示端末を介して取得し、前記空間に関連する特定機器又は特定設備が前記画像センサの視野内に含まれるように前記画像センサを設置する設置ガイダンスを前記表示端末に出力する設置支援部をさらに備える、

請求項 7 記載の行動認識装置。

[請求項9] 前記設置支援部は、前記画像センサが撮影した画像データを取得し、前記画像データに含まれる前記特定機器又は前記特定設備を検出し

、前記特定機器又は前記特定設備の検出結果を示すアノテーション画像を、前記画像データが示す画像に重畳表示させる、

請求項 8 記載の行動認識装置。

[請求項10] 前記設置支援部は、前記アノテーション画像の修正指示を前記表示端末を介して取得し、修正された前記アノテーション画像が示すアノテーション情報をメモリに記憶する、

請求項 9 記載の行動認識装置。

[請求項11] 建物におけるユーザの行動を認識する行動認識方法であって、
コンピュータが、
予め定められた認識対象となる 1 以上の対象行動を含む対象行動情報を取得し、

前記建物の間取り情報を取得し、

前記建物に設置された画像センサの位置情報を取得し、

前記画像センサの位置情報と前記間取り情報とに基づいて前記対象行動情報に含まれる前記 1 以上の対象行動の中から認識候補となる候補行動を選択し、

前記画像センサが検出した画像データを取得し、

第 1 行動認識部が、前記候補行動に応じた 1 以上の認識器を決定し、
前記 1 以上の認識器を用いて前記画像データの特徴量を算出し、

第 2 行動認識部が、前記特徴量に基づいて前記候補行動を認識し、

前記第 2 行動認識部による認識結果を出力する、

行動認識方法。

[請求項12] 建物におけるユーザの行動を認識するためのプログラムであって、
予め定められた認識対象となる 1 以上の対象行動を含む対象行動情報と、前記建物の間取り情報と、前記建物に設置された画像センサの位置情報とを取得する第 1 取得部と、

前記画像センサの位置情報と前記間取り情報とに基づいて前記対象行動情報に含まれる前記 1 以上の対象行動の中から認識候補となる候

補行動を選択する行動選択部と、

前記画像センサが検出した画像データを取得する第2取得部と、

前記候補行動に応じた1以上の認識器を決定し、決定した1以上の認識器を用いて前記画像データの特徴量を算出する第1行動認識部と

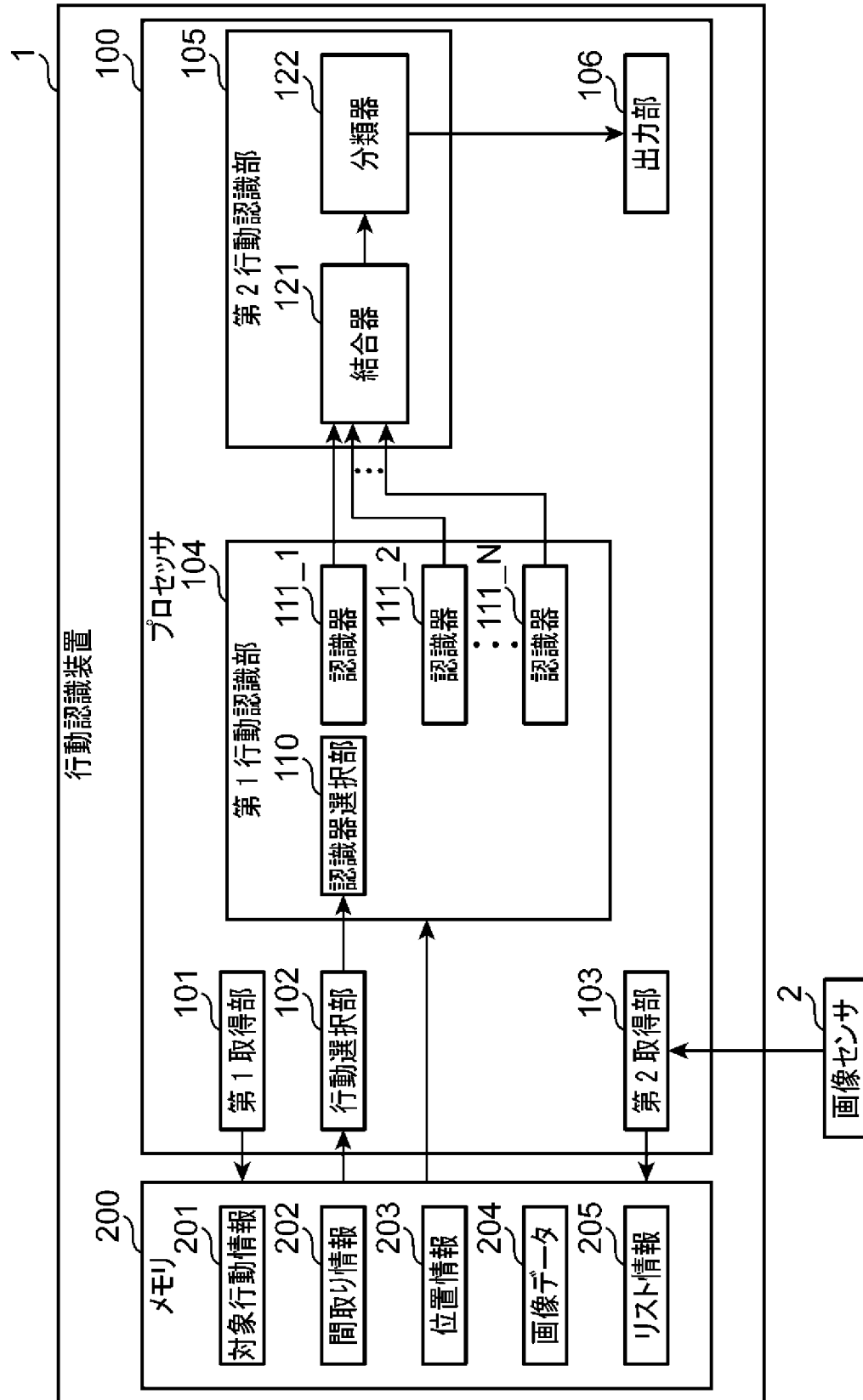
、

前記特徴量に基づいて前記候補行動を認識する第2行動認識部と、

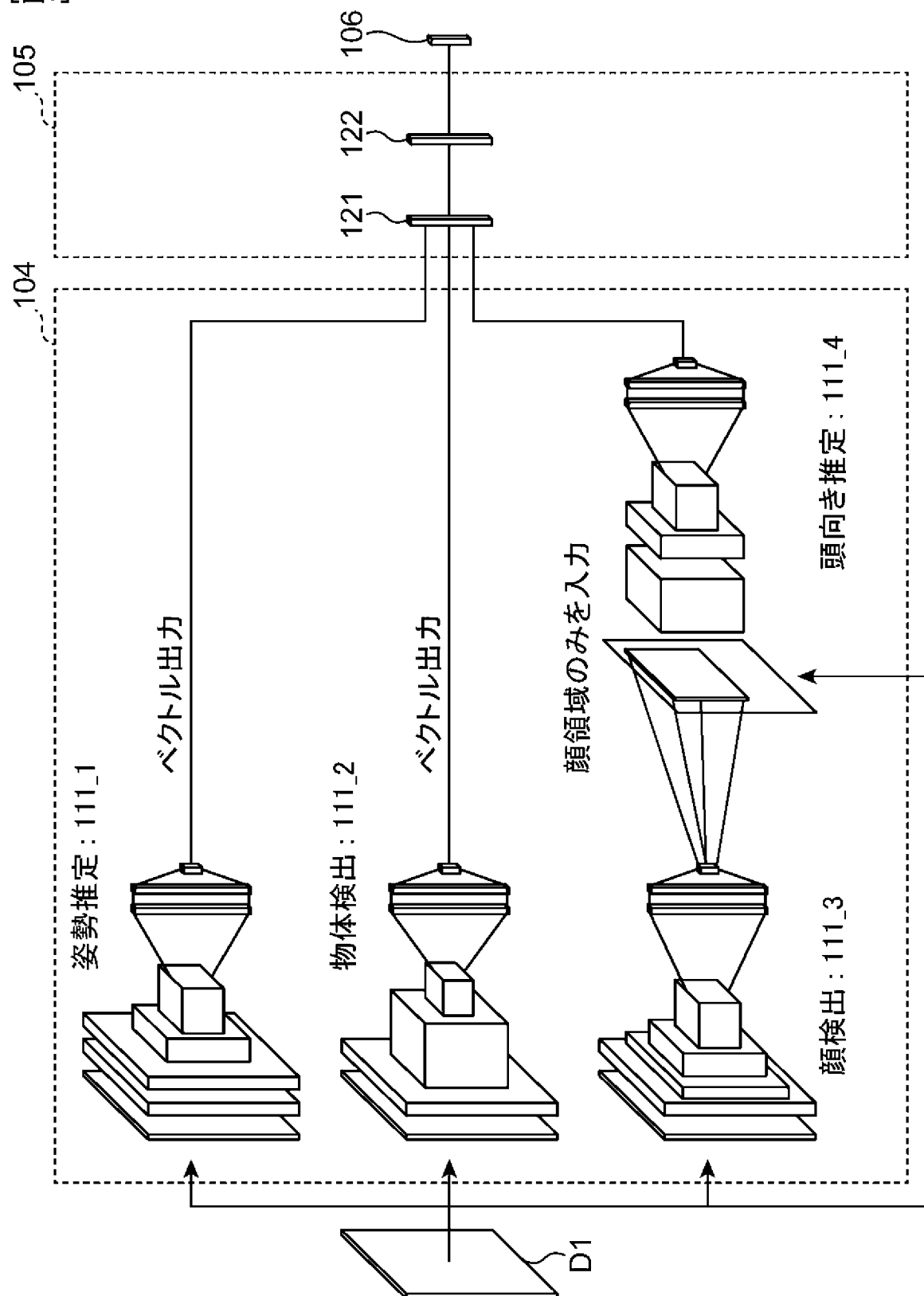
前記第2行動認識部による認識結果を出力する出力部としてコンピュータを機能させる、

プログラム。

[図1]



[図3]



[図4]

T1

対象行動 / 空間	玄関	キッチン	リビング	ダイニング	ベッドルーム	バスルーム	洗面所
拭き掃除	○	○	○	○	○	○	○
掃き掃除	○	○	○	○	○	×	○
歯磨き	×	○	×	×	×	○	○
料理を開始する	×	○	×	×	×	×	×
コーヒーを淹れる	×	○	×	○	×	×	×
ノートPCの使用	×	×	○	○	○	×	×
洗濯	×	×	×	×	×	×	○
読書	×	×	○	○	○	×	×
食事	×	×	○	○	×	×	×
歩く	○	○	○	○	○	○	○

[図5]

T2

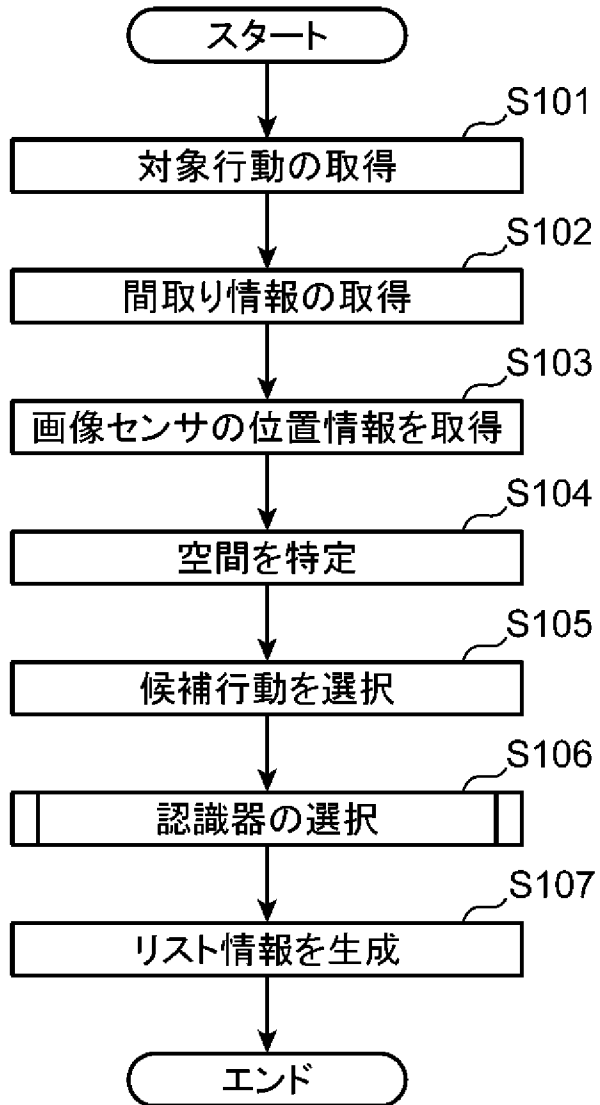
対象行動 / 認識器	姿勢推定	物体検出	顔検出	頭向き推定	年齢性別推定	個人推定	トラッキング
拭き掃除	○	○	×	×	×	×	○
掃き掃除	○	○	×	×	×	×	○
歯磨き	○	○	○	○	×	×	×
料理を開始する	○	○	○	○	○	○	○
コーヒーを淹れる	○	○	○	○	×	×	×
ノートPCの使用	○	○	○	○	○	○	×
洗濯	○	○	○	○	○	○	×
読書	○	○	○	○	×	×	×
食事	○	○	×	×	×	×	×
歩く	○	×	×	×	×	×	○
上記以外の行動を検出する場合	○	○	○	○	○	○	○

[図6]

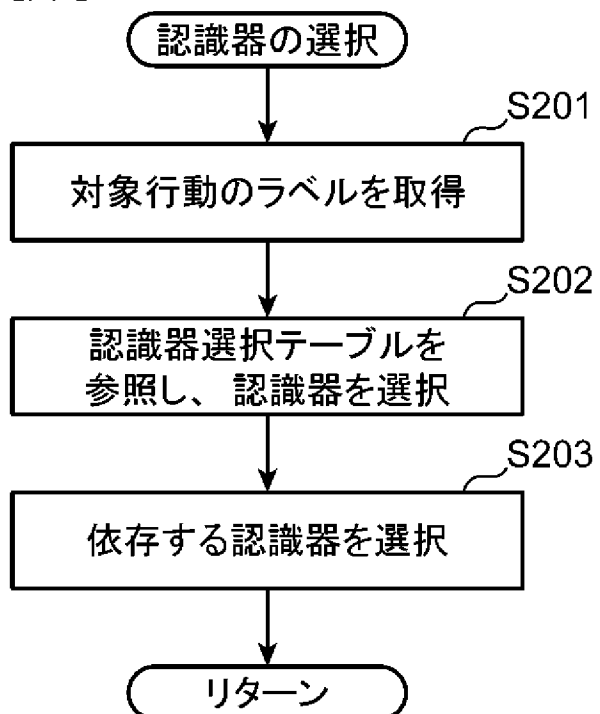
T3

クラス インデックス	ラベル	空間 情報	物体情報	頭向き情報	空間情報の 重み係数 W_r	物体情報の 重み係数 W_e	重み付けの 内容
1	料理を 開始する	キッチン	冷蔵庫、 電子レンジ、 ガスコンロ、 オーブンの 少なくとも 1つが存在	冷蔵庫、 電子レンジ、 ガスコンロ、 オーブンの 少なくとも 1つを向く	キッチンを1、 それ以外は0	冷蔵庫、 電子レンジ、 ガスコンロ、 オーブンは1、 それ以外は0	冷蔵庫、 電子レンジ、 ガスコンロ、 オーブンの いずれかの 方向に頭が 向いた時に W_h を5、 向いていない 場合0、 頭向き以外の 認識器は1
2	拭き掃除	:	:	:	:	:	:
:	:	:	:	:	:	:	:

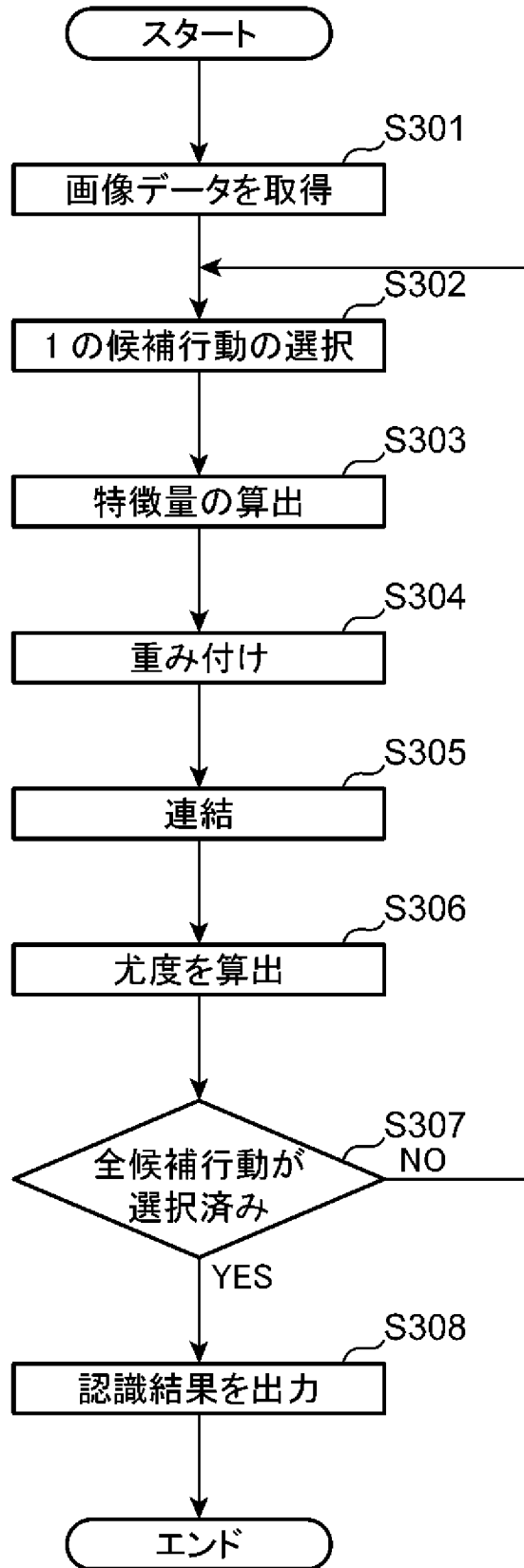
[図7]



[図8]



[図9]



[図10]

T4

識別番号	認識内容	入力 センサデータ	相対計算 コスト	認識器の 依存関係
1	姿勢推定	画像	15	なし
2	物体検出	画像	45	なし
3	顔検出	画像	10	なし
4	頭向き推定	画像	20	3 に依存
5	年齢性別推定	画像	40	3 に依存
6	個人推定	画像	5	3 に依存
7	トラッキング	画像	30	2 に依存
:	:	:	:	:

[図11]

既存の認識器	1フレームあたりの 処理時間 [S]	fps*	ディープニューラル ネットワーク名
A. 姿勢推定	0.0119	84	1 PoseNet
B. 物体検出	0.0151	66	2 SSD
C. 頭向き検出	0.0455	22	3 RetinaFace + 4 DeepHeadPose
D. 行動推定	0.1429	7	5 RepresentationFlowNet
A+B+C	0.0725		

[図12]

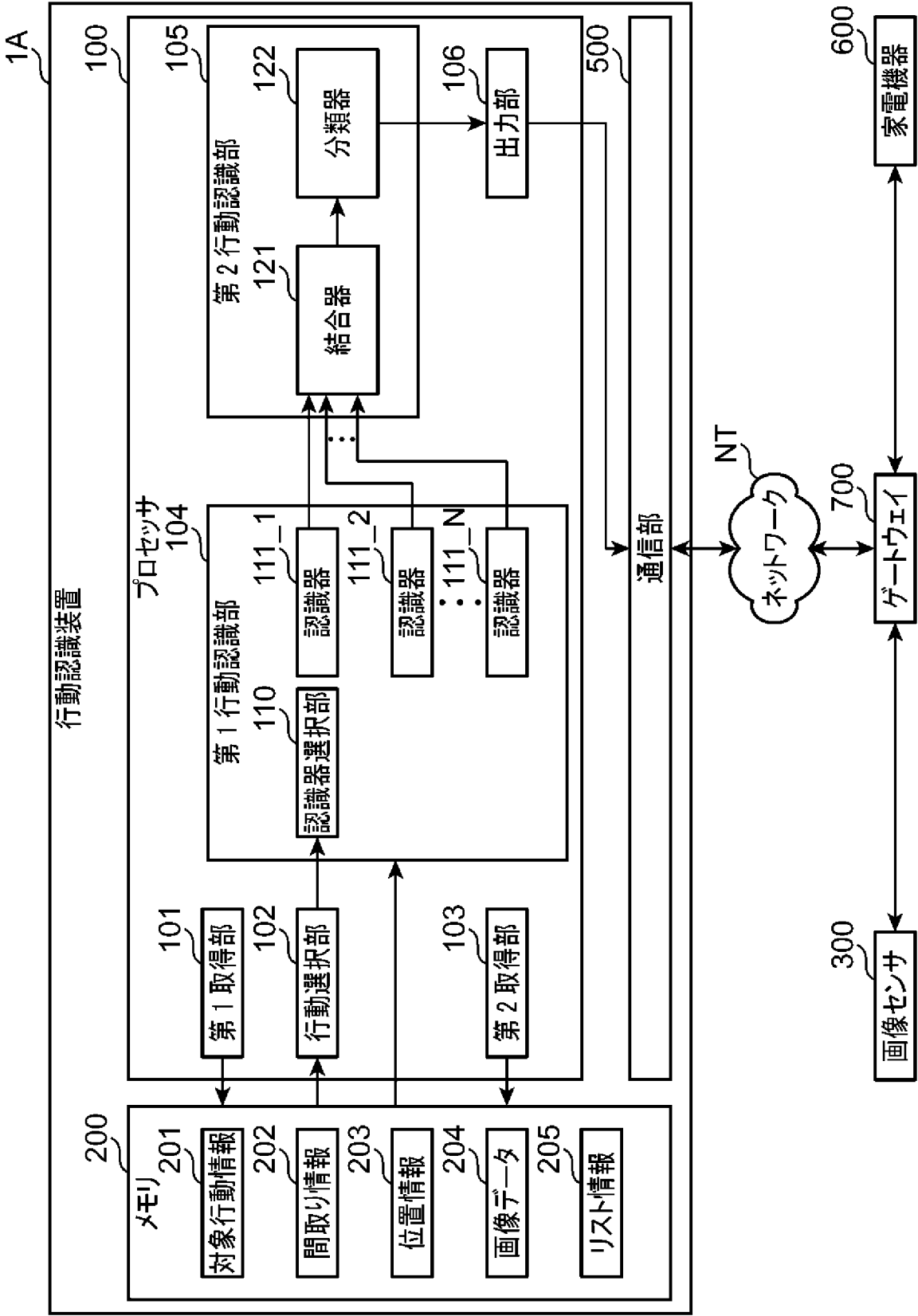
ビデオ ID	「料理を開始する」 のフレーム総数	「料理を開始していない」 のフレーム総数
1	3107	3178
2	1616	1553
3	7762	2302
4	12417	3653
5	1264	2692

[図13]

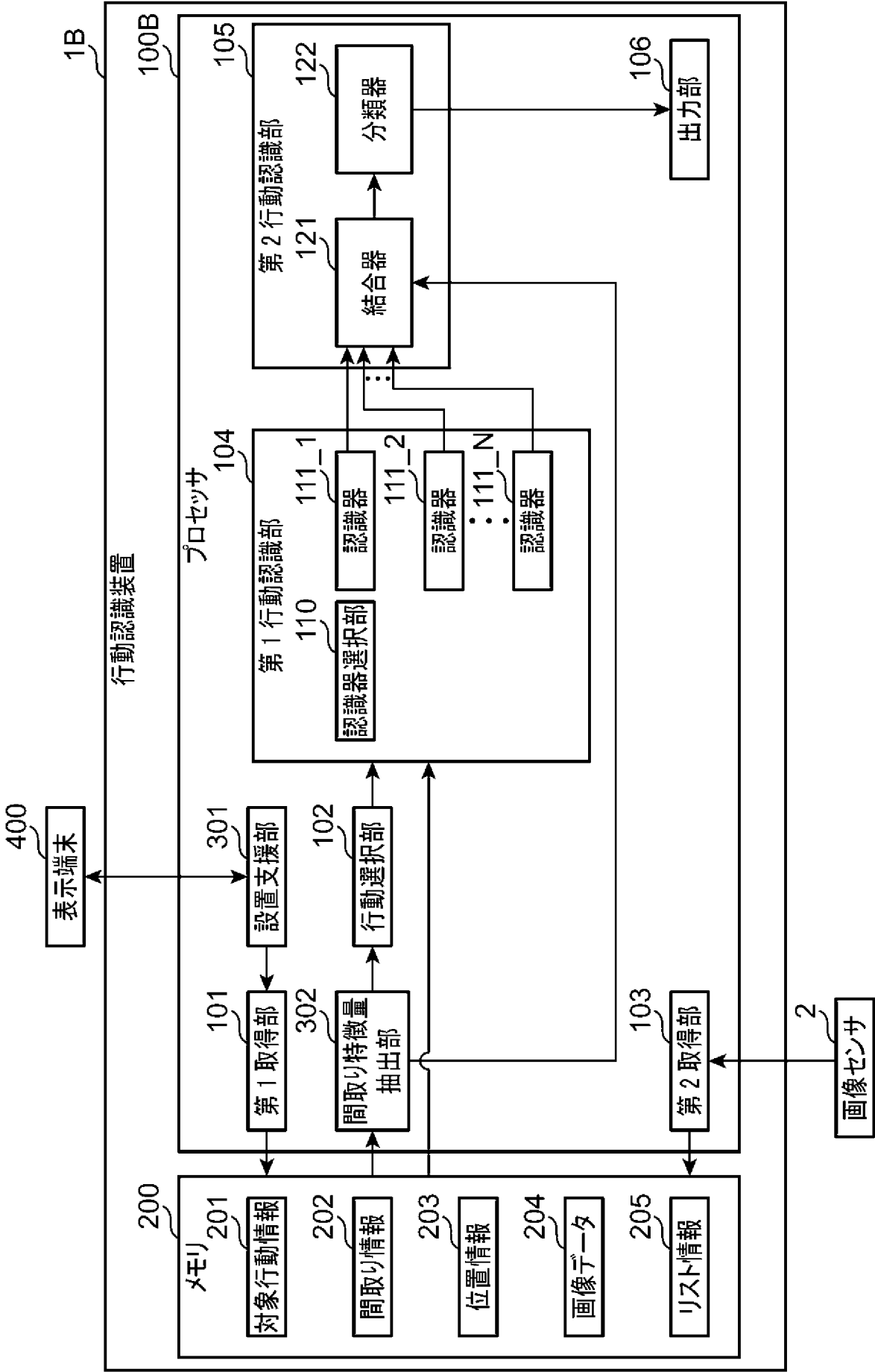
使用した認識器	認識精度
A+B+C	88.5
A+B	88.7
A+C	88.8
B+C	81.6
A	88.7
B	81.6
C	62.6

A: 姿勢推定、B: 物体検出、C: 頭向き推定

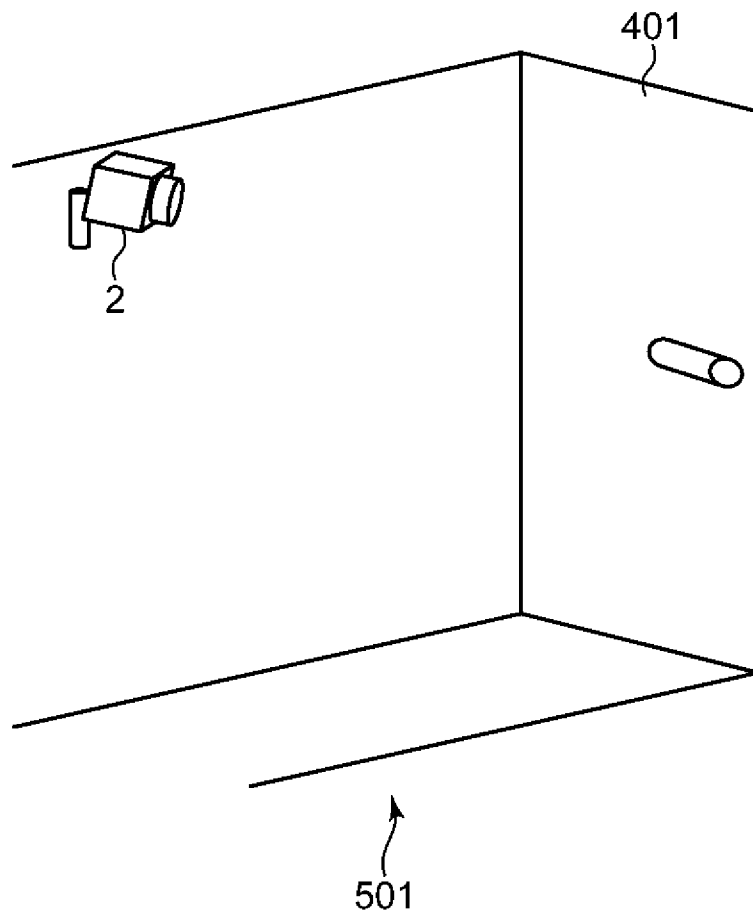
[図14]



[図15]

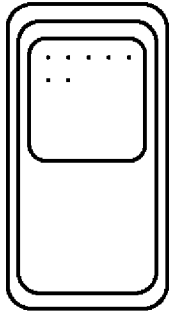


[図16]



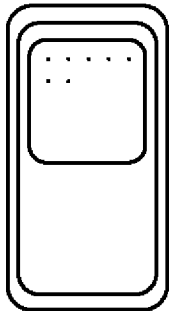
[図17]

400



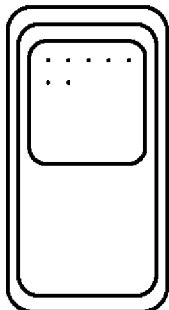
ST1: 使用する場所を選択して下さい。
A: 玄関、B: キッチン、C: リビング ...

ST2: 「A: 玄関」をユーザーが選択



ST3: ドアが映る位置に設置して下さい。
設置後、OK を押して下さい。

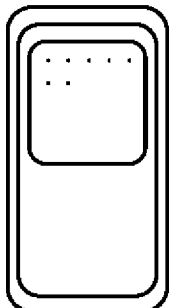
ST4: ユーザーがカメラを設置し「OK」を押す



ST5: ドアを自動検出しています。

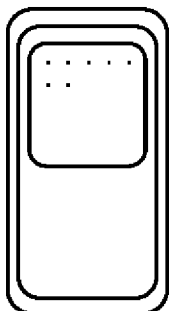
[図18]

400



ST6: 自動検出された赤色ドアガイドの位置を
ドアに合わせるように調整し、OK を押して下さい。

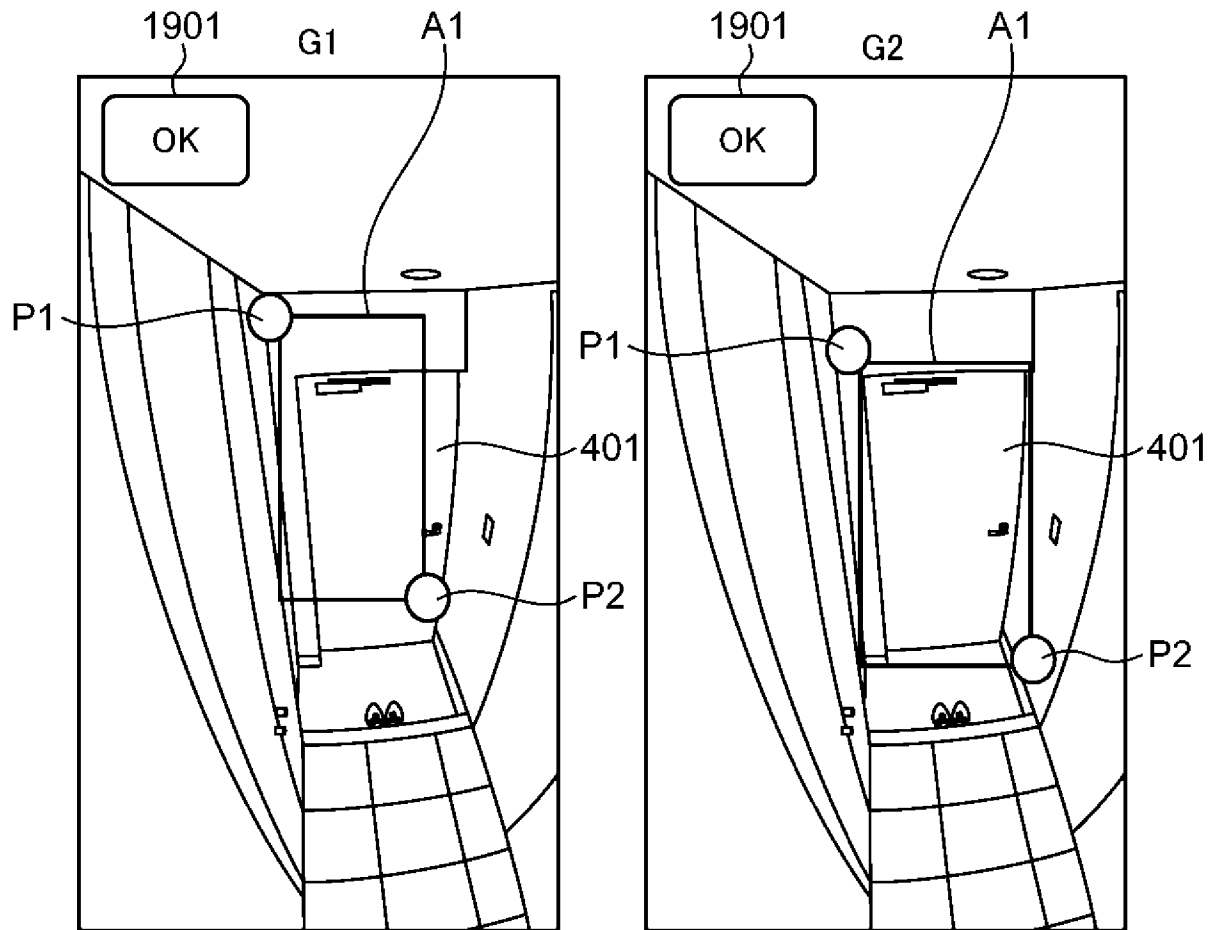
ST7: ユーザーがアノテーション画像を調整し
「OK」を押す



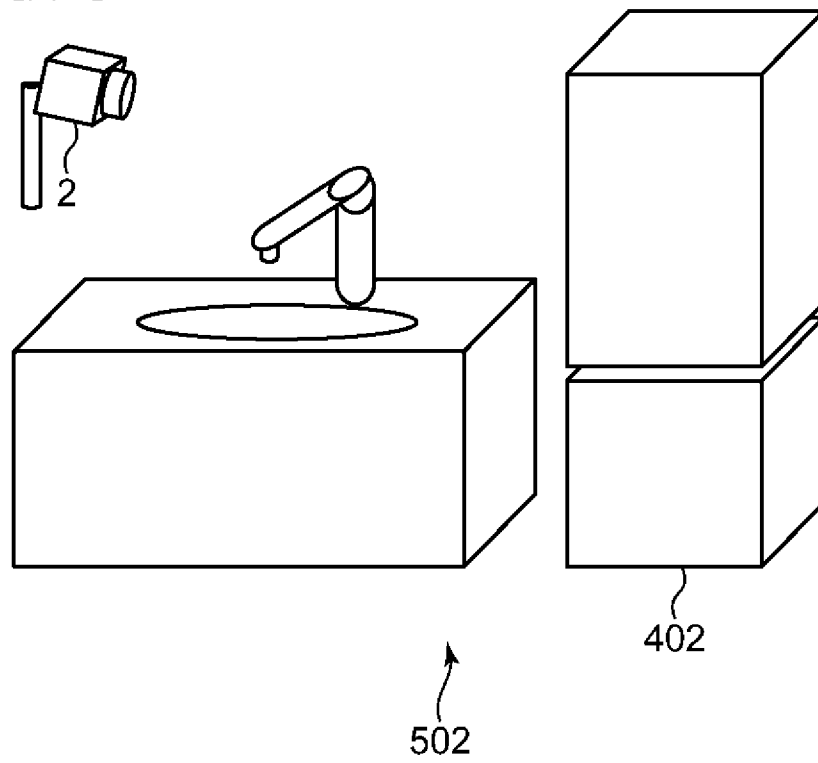
ST8: 設定を保存しました。

ST9: ユーザーが設定画面を閉じる

[図19]

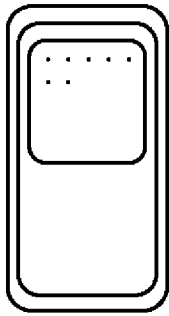


[図20]



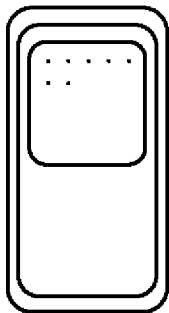
[図21]

400



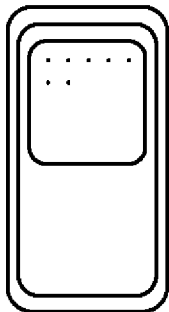
ST1: 使用する場所を選択して下さい。
A: 玄関、B: キッチン、C: リビング ...

ST2: 「B: キッチン」 をユーザーが選択



ST3: 冷蔵庫が映る位置に設置して下さい。
設置後、OK を押して下さい。

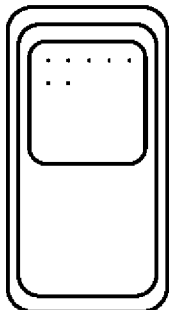
ST4: ユーザーがカメラを設置し「OK」を押す



ST5: 冷蔵庫を自動検出しています。

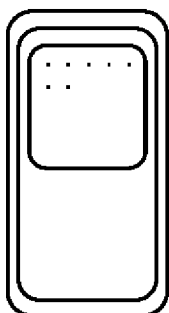
[図22]

400



ST6: 自動検出された赤色ドアガイドの位置を
ドアに合わせるように調整し、OK を押して下さい。

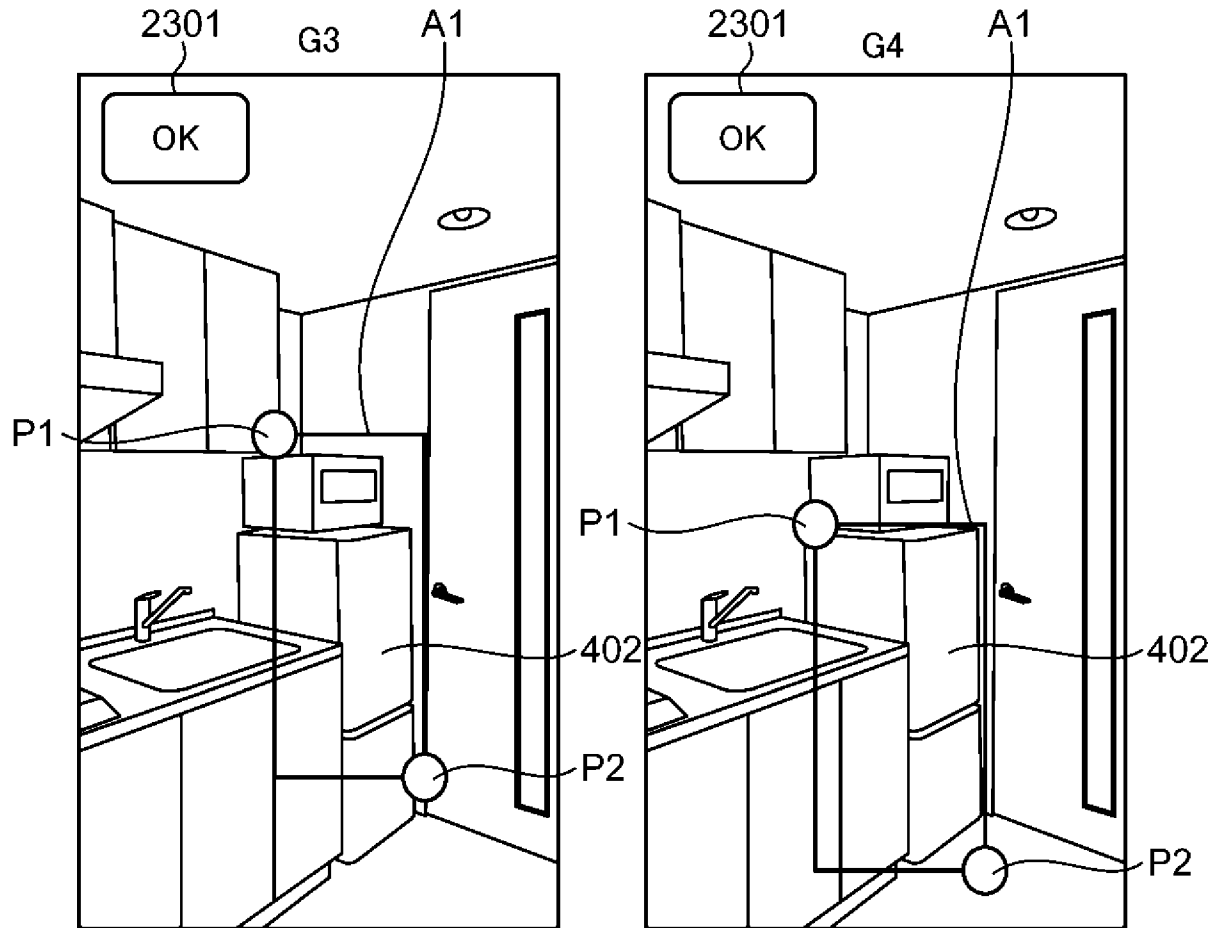
ST7: ユーザーがアノテーション画像を
調整し「OK」を押す



ST8: 設定を保存しました。

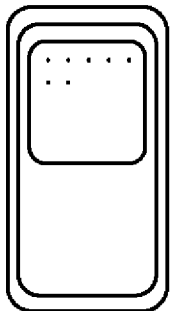
ST9: ユーザーが設定画面を閉じる

[図23]



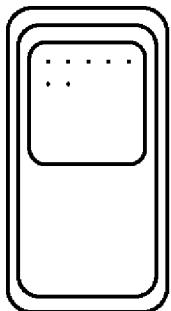
[図24]

400



ST1: カメラ位置がずれてないか確認して下さい。
赤色ガイドが現在の設定です
水色ガイドが自動検出した位置です

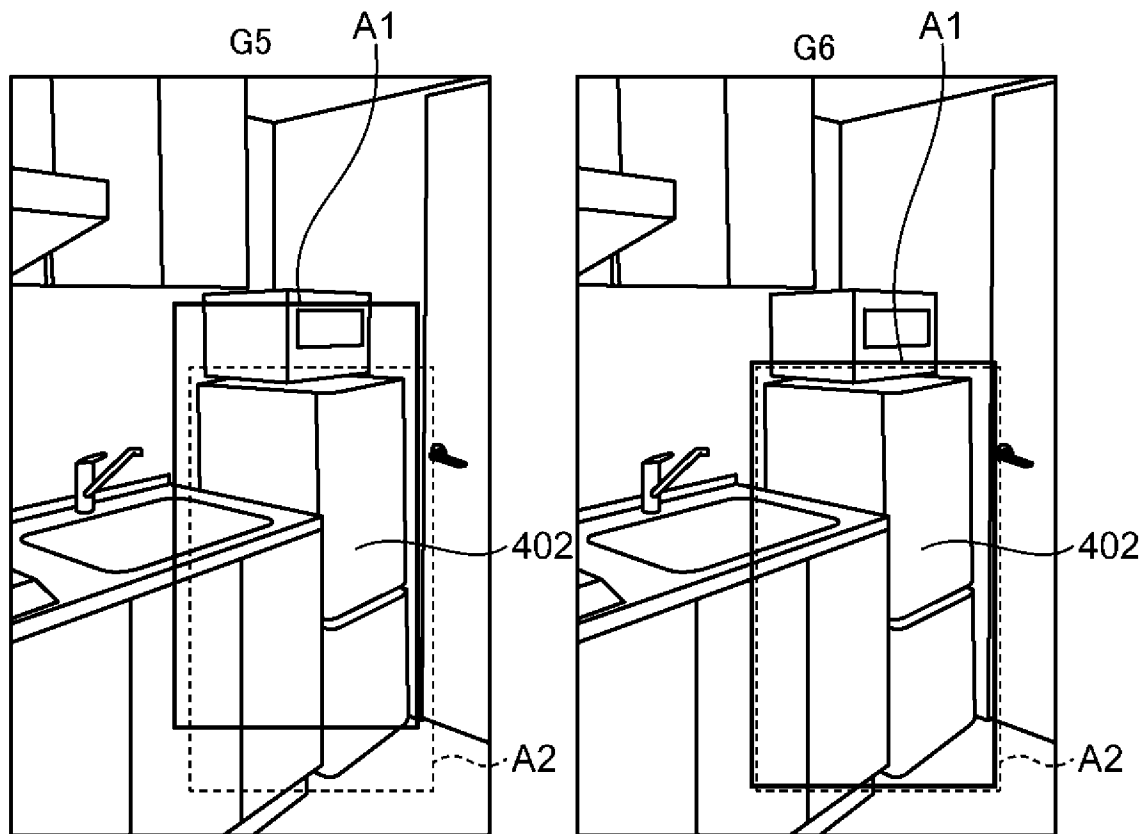
ST2: ユーザーがずれを目視で確認。
設置してあるカメラのアングルを調整



ST3: カメラ位置が正しい位置に戻りました。
(赤色ガイドと水色ガイドが一致した時に
ユーザーに通知)

ST4: ユーザーが設定画面を閉じる

[図25]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/014244

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. G06T7/00 (2017.01) i, G06T7/20 (2017.01) i
FI: G06T7/20 300Z, G06T7/00 660B

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int. Cl. G06T7/00, G06T7/20

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996

Published unexamined utility model applications of Japan 1971-2021

Registered utility model specifications of Japan 1996-2021

Published registered utility model applications of Japan 1994-2021

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	JP 2018-5752 A (HITACHI SYSTEMS, LTD.) 11 January 2018, paragraphs [0006]-[0077]	1-6, 11, 12 7-10
A	JP 2016-38774 A (SHARP CORP.) 22 March 2016, paragraphs [0047]-[0051]	1-12

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search
27.05.2021

Date of mailing of the international search report
08.06.2021

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2021/014244

Patent Documents referred to in the Report	Publication Date	Patent Family	Publication Date
JP 2018-5752 A	11.01.2018	(Family: none)	
JP 2016-38774 A	22.03.2016	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06T 7/00(2017.01)i; G06T 7/20(2017.01)i FI: G06T7/20 300Z; G06T7/00 660B														
B. 調査を行った分野														
調査を行った最小限資料（国際特許分類（IPC）） G06T7/00; G06T7/20														
最小限資料以外の資料で調査を行った分野に含まれるもの <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922 - 1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971 - 2021年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2021年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2021年</td> </tr> </table>			日本国実用新案公報	1922 - 1996年	日本国公開実用新案公報	1971 - 2021年	日本国実用新案登録公報	1996 - 2021年	日本国登録実用新案公報	1994 - 2021年				
日本国実用新案公報	1922 - 1996年													
日本国公開実用新案公報	1971 - 2021年													
日本国実用新案登録公報	1996 - 2021年													
日本国登録実用新案公報	1994 - 2021年													
国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）														
C. 関連すると認められる文献														
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号												
X	JP 2018-5752 A（株式会社日立システムズ）11.01.2018（2018 - 01 - 11） 段落[0006]-[0077]	1-6, 11, 12												
A		7-10												
A	JP 2016-38774 A（シャープ株式会社）22.03.2016（2016 - 03 - 22） 段落[0047]-[0051]	1-12												
☐ C欄の続きにも文献が列挙されている。 ☒ パテントファミリーに関する別紙を参照。														
<table border="0"> <tr> <td>* 引用文献のカテゴリー</td> <td>“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの</td> </tr> <tr> <td>“A” 特に関連のある文献ではなく、一般的技術水準を示すもの</td> <td>“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの</td> </tr> <tr> <td>“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの</td> <td>“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの</td> </tr> <tr> <td>“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）</td> <td>“&” 同一パテントファミリー文献</td> </tr> <tr> <td>“O” 口頭による開示、使用、展示等に言及する文献</td> <td></td> </tr> <tr> <td>“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献</td> <td></td> </tr> </table>			* 引用文献のカテゴリー	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの	“A” 特に関連のある文献ではなく、一般的技術水準を示すもの	“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの	“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの	“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	“&” 同一パテントファミリー文献	“O” 口頭による開示、使用、展示等に言及する文献		“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	
* 引用文献のカテゴリー	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの													
“A” 特に関連のある文献ではなく、一般的技術水準を示すもの	“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの													
“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの													
“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	“&” 同一パテントファミリー文献													
“O” 口頭による開示、使用、展示等に言及する文献														
“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献														
国際調査を完了した日 27.05.2021	国際調査報告の発送日 08.06.2021													
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 岡本 俊威 5H 9178 電話番号 03-3581-1101 内線 3531													

国際調査報告
パテントファミリーに関する情報

国際出願番号

PCT/JP2021/014244

引用文献			公表日	パテントファミリー文献	公表日
JP	2018-5752	A	11.01.2018	(ファミリーなし)	
JP	2016-38774	A	22.03.2016	(ファミリーなし)	